MICHIGAN
ROSS SCHOOL OF BUSINESS

Working Paper

Taxpayer Search for Information: Implications for Rational Attention

Jeffrey L. Hoopes
Department of Accounting & Management Information Systems
Ohio State University

Daniel H. Reck
Department of Economics
University of Michigan

Joel Slemrod
Stephen M. Ross School of Business
University of Michigan

Ross School of Business Working Paper
Working Paper No. 1198
August 2013

TAXPAYER SEARCH FOR INFORMATION: IMPLICATIONS FOR RATIONAL ATTENTION*

Jeffrey Hoopes, Ohio State University, hoopes@fisher.osu.edu
Daniel Reck, University of Michigan, dreck@umich.edu
Joel Slemrod, University of Michigan, jslemrod@umich.edu†

August 30, 2013

Abstract: Prior literature has largely taken taxpayer knowledge of the tax system as given. However, at some point, taxpayers must acquire information in order to respond to, and comply with, tax laws. Using the capital gains tax as a focal case, we examine novel information search data from capital-gains-tax-related Google searches, Wikipedia Page Views, aggregate phone calls to the IRS, aggregate visits to IRS web pages, and aggregate searches conducted on the IRS's webpage. We find strong seasonality in taxpayer information search resulting from a surge in search behavior around tax filing deadlines. This evidence suggests that taxpayers seek information to comply with their tax obligations. We also find evidence that taxpayers search for tax-related information for tax planning purposes. Specifically, we document a year-end ramp-up in capital-gains-tax-related information search, especially searches related to the treatment of capital losses in years following down markets. Capital-gains-tax-related information search co-varies with stock market trading volume, and with Google searches for stock advice, suggesting that taxpayers simultaneously trade and investigate the tax implications of their trading. Finally, we find substantial taxpayer information search around tax events unrelated to either compliance or planning, such as political and news stories related to capital gains taxation. Overall, the data suggest that taxpayers are not fully informed at all times, but that rational attention and exogenous shocks to tax salience drive information search by taxpayers.

JEL Codes: D80, D83, H31, H24

*We wish to thank John Guyton and Ron Hodge of the Research, Analysis, and Statistics Division of the Internal Revenue Service, for helping us use the IRS aggregate administrative data. The views expressed here are those of the authors alone, and do not reflect the views of the Internal Revenue Service. We are grateful for comments on an earlier draft received from participants at the University of Michigan Public Economic Summer Talks series, the University of Uppsala, and at the 2013 Public Economics United Kingdom Conference.

† Corresponding Author:
Joel Slemrod
Ross School of Business
701 Tappan St.
Ann Arbor, MI 48109
Phone: (734) 936-3914
Fax: (734) 936-8716
E-Mail: jslemrod@umich.edu

1. Introduction

Taxes, especially income taxes, can be complex and confusing. Despite a general awareness of this fact, the consequences of complexity and misunderstanding are not well-understood. Survey evidence suggests that many taxpayers do not understand basic tax concepts (Blendon et al, 2003), and the compliance cost of taxes, including learning enough about them to be able to comply, is large (Slemrod, 1995 2004). Given confusion surrounding tax incentives and tax law, it seems plausible that many taxpayers simply ignore, or misperceive, the incentives built into the tax code when they make decisions with tax consequences. Alternatively, taxpayers might collect information on the taxes they suspect matter to them, and use this information to make tax-efficient decisions. Finally, taxpayers might learn just enough about tax policies to enable them to fill out their tax return and perhaps avoid an audit.

In this paper, we investigate these alternative views of taxpayer information search. We are able to reject that taxpayers are fully informed about the tax system, that they act in complete ignorance of the tax system, and that they gather information only for tax compliance purposes. The evidence we present suggests instead that taxpayers employ rational attention to tax policies, in line with theories proposed by Sims (2003) and Reis (2006). In addition, we observe that exogenous shocks to tax salience from news events can substantially increase information search.

Modern technology has greatly expanded the accessibility of information. Any person with access to the Internet may, in a few minutes, learn at least something about the most obscure details of the tax code, and any person with a phone can call the IRS and ask questions about his or her specific tax situation. People undoubtedly do use these resources to seek information:

people Google "tax" more often than they Google the names of public figures,[1] the IRS website has received on average 4.6 million visits *per day* since 2004 and, before it was largely supplanted by the website, the IRS call line would receive hundreds of calls per day on capital gains taxes alone. How tax knowledge matters hinges on 1) how and when people seek out tax-related information, and 2) whether they change their behavior once they acquire it. In this paper, we address primarily the first of these questions, and provide some preliminary analysis of the second.

This paper utilizes high-frequency aggregate data from IRS administrative records and publicly available sources to study information-seeking about taxation in order to understand better the links between government policy and taxpayer behavior. We study when and through what channels people seek information about important tax policies, and we explore the effect of several events on information-seeking and tax-related behavior. Our data consist of aggregate high-frequency time series on calls to the IRS toll-free phone number, aggregate visits to the IRS website, measures of Google searches on tax-related terms, and views of tax-related web pages on Wikipedia. We also use supplementary data on searches conducted on the IRS website and referring pages to the IRS website. From these sources we collect data on information seeking regarding one specific tax-related topic: capital gains taxes. We select this topic because it is a perennially controversial policy issue, because data on the relevant taxed behavior, sales of capital assets, are available on a high-frequency basis, and because the American Taxpayer Relief Act of 2013 (ATRA) enacted a change in capital gains tax rates during our sample period.

In this analysis we will take as given the supply of information, and study changes in the demand for information. Information content is made available by the IRS, by accounting firms

[1] For example, Google searches for "tax" were more common than searches for "Barack Obama," and were more common than "Obama" at all periods except the weeks of the Presidential Elections of 2008 and 2012, when Barack Obama was elected. This claim may be verified using publicly available data at http://www.google.com/trends/.

and other organizations in the business of providing information generally, such as Wikipedia. The IRS-provided information can be accessed through the Internet, via printed information booklets, and through toll-free numbers. Person-specific information can also be obtained by perusing one's paycheck, one's Form W-2, or one's prior tax returns. Both information from the IRS website and information from non-government organizations may be found quickly using Google searches. In all cases, the marginal cost of information is only the time spent acquiring it, provided one has an Internet or phone connection. The time cost of tax information may have decreased with the advent of the Internet, and search engines such as Google in particular. Long-run trends in information search may therefore vary due to both the supply and demand of information. By focusing on high-frequency variation in information search, we isolate variation in the demand for information, since the supply changes very little from day to day.

We will use data on information search to distinguish between competing theories of information search, which correspond to different assumptions on the role of information in decision-making. We study information-seeking around five different types of events: 1) time notches, 2) macroeconomic changes, 3) policy changes or the mention of potential policy changes, 4) filing deadlines or approaching filing deadlines, and 5) tax-related news events.

Information search around the first three types of event should be tied to the taxpayer's ability to make fully informed decisions affected by tax incentives. First, a particular date is often relevant for the incursion of tax liability or the tax efficiency of behavior. We call these time notches to indicate that tax liability can change abruptly, and non-discretely, at certain dates, usually at year-ends. For example, the last date to affect one's capital gains tax liability in a given year is December 31st of that year. Therefore, as the deadline approaches individuals may research the benefits from realizing a capital gain or loss in the current year as opposed to a future year. Second, macroeconomic changes may be associated with information search due to perceived changes in the importance of understanding the tax consequences of a particular behavior. For example, asset owners may seek information about capital gains taxation as the stock market collapses to understand the tax implications of their mounting losses. Third, actual changes or the mention of possible changes in tax policy or enforcement policy may cause individuals to learn about the tax system, either to plan for the future or to make an informed voting choice.

Taxpayers might also seek information on the tax system when completing their tax return. As such, the fourth set of events we study are *filing deadlines* and the onset of these deadlines, when individuals may learn about taxes due to approaching deadlines for filing tax returns. For the taxes we consider, the deadline affecting most taxpayers is in mid-April.

News events sometimes spark public interest in tax policy, and in this case individuals may search for tax information in order to develop an informed opinion on current events, or perhaps merely out of curiosity. The fifth kind of event we study consists of news stories such as the release of a public person's personal tax information. We can view these events as exogenous shocks to the salience, or visibility, of a particular aspect of tax policy. They have no direct bearing on an individual's compliance responsibilities, nor do they directly affect tax-related incentives. The search may, however, provide information that does inform the searcher about her own current or future decisions.

We observe strong seasonality in the search for information on capital gains taxes through all channels. The form of this seasonality indicates that taxpayer information search increases substantially during the period commonly called "tax season," which runs from mid-January to mid-April of each year. An even more pronounced spike in information search occurs very close to the filing deadline in mid-April. The data suggest that aggregate calls to the IRS hotline are almost completely explained by search behavior during tax season, while Google and Wikipedia are sources of information for purposes other than tax compliance alone, such as tax planning.

We document the impact of several discrete events on taxpayer information search regarding capital gains taxes, through Google and Wikipedia. Presidential debates in which candidates discuss their proposals for capital gains taxes, presidential elections, and policy changes all generate large and significant increases in taxpayer information search. In every case, these events cause a spike in taxpayer information search that fades within three to four days. The passage on January 2, 2013 of ATRA, which increased the top marginal tax rate on capital gains, generated the largest spike in information search in the Google and Wikipedia data. Surprisingly, the release of then presidential candidate Mitt Romney's 2010 tax return in January of 2012 also caused a large spike in information search, second in size only to the passage of ATRA. Mr. Romney's low effective tax rate generated substantial news coverage. Apparently, many Americans then researched the principal cause of the candidate's low tax rate, which was that capital gains—a major source of Romney's income—are generally taxed at a lower rate than labor income.

Next, we examine more closely the link between information search and behavior. Macroeconomic change also affects information search through Google and Wikipedia. On days with large trading volume in the stock market, we observe significantly elevated information search related to capital gains taxes. Notably, the elevated information search occurs only within a single day: trading volume is not significantly related to prior or subsequent information search. We also explore a measure of information search for personal investment advice generally, using Google searches. Searches for such advice directly indicate individuals' information seeking when contemplating a decision which has tax consequences. Daily searches for stock advice are predictive of searches for capital gains tax on the same day, the previous day, and, in some specifications, one day in the future. This is the first evidence that some taxpayers investigate the tax consequences of an action while contemplating the action itself. Notably, searches for stock advice are even a better predictor of information search for capital gains taxes than stock market trading volume.

These patterns indicate that information search is consistent with a model in which individuals search when different events make understanding tax policy more important, i.e. when they increase the return to information search. However, an alternative explanation might be that individuals make decisions with tax consequences and *then* research the tax implications of those decisions. For example, an individual might sell stock and realize a capital gain without considering the tax effect, and then research capital gains taxes shortly afterwards in order to understand the effect on his wealth of the stock sale. Similarly, policy changes might lead to research for the purpose of becoming an informed voter, but not for making behavior tax-efficient. We next provide evidence that more strongly suggests a causal relationship, whereby events cause individuals to search for information in order to make more informed decisions. We do this by examining information search on capital losses around the 2008 stock market crash.

We document substantial elevation in information search related to capital *losses* near the end of 2008. We argue that information search near the year-end time notch that is isolated in information about capital losses (as opposed to gains) is consistent with taxpayers rebalancing portfolios in order to take tax-efficient advantage of capital losses resulting from the 2008 market downturn. If taxpayers merely wanted to know the effect of losses on their after-tax wealth, they would have searched as the losses were occurring (which they did to some degree). And, if taxpayers merely sought to understand the treatment of capital losses in order to comply with their tax filing obligations, we would only see an increase in information search about capital losses around the filing deadline in 2009 (which we also observe). Observing increases in information search 1) as the losses occurred, 2) prior to the time notch and 3) during tax season in 2009 is consistent with taxpayers 1) trying to understand the impact of events as they occur, 2) using knowledge of the tax system to affect their after-tax wealth, and 3) seeking information to comply with their tax filing obligation. These patterns are consistent with information search conforming to a rational attention model of information, wherein taxpayers seek information when the gains to doing so are sufficiently large.

The different reasons that individuals search for information are potentially inter-related in interesting and policy-relevant ways. We identify some notable interactions between different patterns in information search. In particular, we show that stock market activity throughout the year is associated with elevated search behavior around both time notches and the following tax season, through Google searches, Wikipedia, and (for tax season only) calls to the IRS. More specifically, in years of large market movements—i.e. years with large trading volume or large annual market returns—we observe elevated search volume on capital gains taxes at year end, and during the subsequent tax season.[2]

Finally, we take a broader focus to analyze use of the IRS website and call line for individuals. Overall use of information provided by the IRS, as measured by the number of daily page views and visits to the website and calls to the call line for individuals, is primarily driven by tax season, as with measures of information search about capital gains specifically. One exception to the seasonal pattern occurred in May 2008, when the stimulus checks distributed

[2] These interactions are of course only suggestive: many unobserved events could happen in the same year as the year of a policy change or a year with high sales volume, which might also affect information search in the subsequent year.

through the tax system generated a large spike in visits to the IRS website and calls to the call line.[3] Millions of taxpayers apparently sought information in an attempt to know whether they would receive a stimulus check from the government and, if so when it might come and how large it might be.[4]

Our attempt to make inference about different types of learning from patterns in information search and filing behavior is subject to the limitation that we do not observe all channels of information acquisition. For example, we cannot observe taxpayers obtaining information through their financial advisors (although our measures may capture their financial advisors' information search). Further, we cannot rule out that individuals "learn by doing," whereby individuals learn about tax incentives while filling out a tax return and then adjust their future behavior based on what they learned. Nor can we observe learning through social networks or other channels (as Alstadsæter, Kopczuk, and Telle (2010) suggest does occur). As in all similar studies, we also cannot quantify the *amount* of learning that occurs via acquiring any particular piece of information.

The rest of the paper is organized as follows: Section 2 reviews relevant parts of the literature, including models of information and attention. Section 3 summarizes capital gains tax law and tax incentives. Section 4 describes the data in detail. Section 5 describes the research design. Section 6 provides and exploratory analysis of the raw data. Section 7 estimates the impact of discrete events. Section 8 provides evidence suggesting a causal relationship between

[3] We know that the spike in May 2008 resulted from the stimulus and not some other simultaneous event by noting that "stimulus" and "stimulus check" were among the five most popular searches performed on the search bar on the home page of the IRS website.

[4] If we assume that attention to the stimulus was rational, this result further suggests that many taxpayers' utility must have been increased significantly from knowing the timing and amount of their stimulus checks, which speaks to the efficacy of the stimulus. Wanting information about the exact date of arrival of a stimulus check makes sense only to someone not planning to simply save the money.

macroeconomic change and information search, Section 9 examines patterns in the use of the IRS website and call line, and Section 10 concludes.

2 Literature Review

This section briefly reviews prior literature on information seeking, both in the public finance literature related to taxes, and other literatures. Throughout this paper, we adopt a notion of information-seeking similar to that articulated by Marchionni (1995, pp. 5-6), "a process in which humans purposefully engage in order to change their state of knowledge" or, as in Zerbinos (1990, p. 922), "when a person recognizes a gap in their knowledge that may motivate that person to acquire new information."

2.1 Macroeconomics and finance

Information plays a large role in the modern macroeconomics literature. Many macroeconomists have considered how best to account for the inertia in observed economic behavior and have addressed to what extent it can be accounted for by imperfect information. For example, the models in Sims (2003) are motivated by the idea that information that is freely available to an individual may not be used, because of the individual's limited information processing capacity. Alternatively, in Reis (2006) consumers rationally choose to only sporadically update their information and re-compute their optimal plans, while in between updating dates they remain inattentive. The optimal length of inattentiveness weighs the costs of reacting with a delay to news against the costs incurred by planning. Agents with rational inattention respond with more delay and information-processing error — or may not respond at all — to fluctuations that are small and therefore relatively unimportant to them. Both models imply that news disperses slowly throughout the population, so events have a gradual and delayed effect on behavior.

Financial economics research has long struggled with how to account for the acquisition of knowledge. As Grossman and Stiglitz (1980) suggest, when knowledge about asset prices is obtained only at a cost (as is the case in reality), market equilibria that require perfect knowledge may not obtain. A recent literature in financial economics has taken advantage of newly available data and focused on the demand by investors for information. Da, Engelberg, and Gao (2011) propose the Google Search Volume Index (GSVI) as a direct measure of investor attention. Drake, Roulstone and Thornock (2012) use the GSVI for public company ticker symbols to examine the timing and magnitude of Internet search around earnings announcements and the factors that influence Internet search.[5] To understand the determinants of search volume, they regress the GSVI on a broad set of explanatory variables that fall into one of five classifications: event dates, media attention, news, liquidity, and distraction; among their results of interest is that investors extend more effort when the potential returns to search are higher. Choi and Varian (2009, 2012) argue that aggregate search frequency is a direct and unambiguous measure of attention and provide evidence that it can predict home sales, unemployment claims, automotive sales, and tourism.

2.2 Public economics

Political science research is fairly persuasive that voters know very little about the details of government. Delli Carpini and Keeter (1996), in a comprehensive survey of the political knowledge of voters covering several decades and hundreds of surveys, show that the majority of voters are ignorant of many key aspects of the U.S. political system, such as who has the power to declare war, the respective functions of the three branches of government, and who controls

[5] In another context, Ginsberg et al (2009) find that search terms related to influenza can predict flu outbreaks one to two weeks before Centers for Disease Control. Mestyán et al (2012) find that Wikipedia page view data suggestive of interest in upcoming releases of new movies can predict box office success.

monetary policy. The comprehensive survey done in 2003 by the Kaiser Foundation, the Kennedy School of Government, and National Public Radio confirms that this statement is true as it applies to taxation (Blendon et al (2003) summarize the survey results).

Accurately assessing one's income tax liability, much less one's marginal tax rate, is difficult for many taxpayers. To be sure, this ability is heterogeneous because people differ in their cognitive ability and because the cognitive process requirements of tax compliance vary greatly across taxpayers. For those with complicated financial affairs, especially with respect to capital income, it can be very difficult. For this reason, many taxpayers purchase professional and software assistance, which can facilitate compliance with tax law but may also reduce taxpayer familiarity with the details and implications of the tax code (i.e., Goolsbee, 2004).

One specific setting where researchers have investigated understanding of the tax system is taxpayers' perceptions of their average and marginal tax rate. Brown (1968) and Fujii and Hawley (1988) find that individuals' self-reported marginal tax rate often differs from the true rate that can be estimated from their demographic characteristics. Research by de Bartoleme (1995) shows that, in a lab experiment, MBA students often confuse the average tax rate with the marginal tax rate when making investments in a taxable versus non-taxable project. Sheffrin (1994) reviews studies of American, British, and Canadian taxpayers that find that taxpayers generally underestimate both their total tax liabilities and their marginal tax rates. Liebman and Zeckhauser (2004) suggest that, because of cognitive limitations, taxpayers presume that their marginal tax rate is the easier-to-calculate average tax rate; they call this rule of thumb behavior "ironing," one of two examples of what they dub "schmeduling," defined as an inaccurately perceived price schedule. Finally, substantial unsystematic evidence from teaching undergraduate and master's level students suggests that over half of these students begin class mistakenly believing that the kinks in the US income tax rate schedule are in fact notches, so that reporting one more dollar of income triggers hundreds of dollars of additional tax liability.[6]

That taxpayers have cognitive limitations has many implications for tax analysis. One is that the distribution of tax burden may depend on, in addition to the intended characteristics, cognitive ability. Another is that taxpayers who are not so good at addressing tax matters (or are unwilling to take illegal advantage of the tax system) may avoid certain employment statuses, such as self-employment, that require or reward this kind of savvy.

Recently, the public economics literature has settled on the term “salience” to capture the extent to which tax aspects of the environment are noticed, and acted upon, by those affected. A seminal paper in the modern literature is Chetty, Looney, and Kroft (2009), who in a field experiment find that posting at the store shelf retail-sales-tax-inclusive prices, rather than adding taxes at the point of sale, reduces purchases. Their findings comes from an experiment at a large grocery retailer in California, at which prices inclusive of the 7.375 percent state sales tax were posted alongside the original pre-tax price over a three-week period for three product categories (cosmetics, hair care accessories, and deodorants). Using other products in the same store and two neighboring stores as experimental controls, they estimate that the “tax treatment” reduced demand by 8 percent; given demand elasticities of 1 to 1.5 for the affected products, they conclude that most consumers do not take the sales tax revealed at the cash register into account.

2.3 Testable implications of theories

Although our analysis focuses on taxpayers’ search for information about income taxes, it has more general implications regarding the information economic agents possess when making decisions (in our case, often very economically large decisions). Do individuals possess full

[6] After repeatedly correcting this mistaken belief, one can by the end of the course get the fraction of students subject to this misperception down to about one-third.

information for important economic decisions? If not, is attention to important information rational? If attention is rational, how should we model the information acquisition process? We will shed some light on the answers to these questions using patterns in the timing of information search. Table 1 summarizes the predictions of the theories our analysis illuminates, as described in this section.

The null hypothesis in our analysis is that taxpayer information search is unresponsive to political or economic events. This hypothesis would be implied by the assumption of full information, wherein an individual always has all data necessary to make an optimal choice. Although frequently criticized and relaxed in the economics literature, this assumption is still common in public finance for modeling the response to tax incentives and the welfare cost of taxation. If individuals have full information, it stands to reason that they should not search for information. Why spend time researching something that is already completely understood?

In rational expectations models like the one proposed by Muth (1961), the individual faces uncertainty about the future, but knows everything about the present and costlessly and instantaneously absorbs all new information. Rational expectations might allow searches for tax information to respond instantaneously to unanticipated policy changes or shocks to the probability of a policy change, as these require that the individual absorb new information about policy. However, information search should not respond under rational expectations to economic or political events which change the importance of understanding tax policy, but not tax policy itself. Information search should also be unresponsive to the implementation of policy changes that were public prior to implementation. Individuals should already possess all publicly available information on tax policy.

An attractive alternative is the concept of rational attention, under which individuals have limited capacity for processing new information, and they allocate that resource optimally. To the best of our knowledge, all previous theoretical and empirical work on rational attention has assumed that attention itself is unobservable, but we take the approach of directly observing attention and testing whether it conforms to the assumptions embedded in the theory. The economics literature has not yet settled on a canonical model of rational attention. Alternative assumptions for how limited attention may be rationally allocated are proposed by Sims (2003)[7] and Reis (2006).[8] In either model, information search responds positively to events that increase the utility gain to understanding the incentives at play in making a decision (such as a stock market downturn or approaching time notch) and to events which introduce new information about incentives (such as policy changes).

One difference between these two models in the literature concerns the precise timing of information search. In Sims' (2003) approach, agents constantly update their beliefs based on their attention to different sources of information, so we should expect information search to instantly react to events that change the importance of understanding a given concept. In Reis' (2006) approach, updating beliefs is costly and upon updating the individual simultaneously plans future consumption and decides when to update again. In between updating dates, the individual is inattentive to new information. Events like policy changes should affect information search in the Reis (2006) model, but the searching should be spread out over a longer time period as not all individuals constantly update beliefs. We can therefore use the timing of information

[7] See Sims (2006) and Maćkowiak and Wiederholt (2009) for work building on the Sims (2003) approach to rational attention.

[8] For an application of the rational attention approach to taxes specifically, see Chetty, Looney, and Kroft (2007). These authors pose a static model of rational attention to tax incentives, in which the individual must either ignore a tax or pay some cost to attend to the after-tax price. Because their model is not dynamic, it does not carry a clear implication for the timing of information search.

search around unanticipated economic and political events to test the difference between the mechanisms for rational attention employed by the two models. A near instantaneous response indicates that the data favor the Sims (2003) mechanism over the Reis (2006) one.

Attention is also the subject of a large literature in psychology. One useful distinction in this literature is between "exogenous," or "stimulus-driven," attention and "endogenous," or "goal-directed," attention.[9] The concept of endogenous attention is essentially the same as rational attention: the individual voluntarily directs attention to meet some goal. Under exogenous attention, though, the individual's attention and search for information are driven not by the belief that she will gain utility from new information, but rather by an external stimulus,[10] such as increased salience. In this version of events, the individual might search for information on Google regarding anything she sees in the news which she does not understand, regardless of her own utility gain from information search. If attention to economic information is primarily exogenous,[11] information search should respond to news events that mention policy—because these are usually accompanied by news coverage which makes the policy more salient—but not to other events which make understanding policy important.

3 Capital Gains Taxation and Behavior

In this paper we focus on one aspect of the US income tax—the taxation of capital gains. In this section we provide some background to this topic.

3.1 The Tax Law

[9] For more on this literature, including the neurological foundations of both types of attention, see Theeuwes (1991), Yantis (2001), and Connor, Egeth, and Yantis (2004).

[10] The experimental psychology literature has in mind different stimuli than we examine here, but the spirit of the exogenous-endogenous distinction is the same.

[11] The psychology literature views the exogenous-endogenous distinction as complementary explanations of attention, rather than competing ones. For our purposes, the question is which types of attention are important for attention to economic information.

Capital gains generated from the sale of capital assets have received preferential treatment relative to labor income since 1921 in the United States. Income from the sale of capital assets (e.g., corporate stock or bonds) is recognized in the year of sale, and the taxable income is equal to the sale price of the asset less its tax basis (the historical price plus any acquisition costs and improvements to the asset, less any accumulated depreciation).[12]

Capital gains and losses are divided into two categories, short-term and long-term. Long-term capital gains and losses arise from the sale of a capital asset that has been held for more than one year, and short-term capital gains and losses arise from the sale of assets held less than one year. Long-term capital gains receive a favorable tax treatment, and are currently taxed at a maximum of 20%.[13] Short-term gains are taxed as ordinary income.

Gains and losses from short and long term assets are netted together in well-defined ways to determine a taxpayer's ultimate tax liability. A limit of $3,000 of net capital losses (capital losses in excess of capital gains) may be deducted by individual taxpayers per year against ordinary (i.e. non-capital-gains) income, but can be netted against capital gains. If a taxpayer dies without "recognizing" (usually by sale) the gain, the gain goes untaxed, as the inheritor receives the capital assets with the tax basis equal to the market value of the asset on the date of the death of the taxpayer (a "step-up" in basis). This transfer of assets will still be subject to the estate tax, although estate tax planning may help mitigate the effect of the estate tax.[14]

Because capital gains taxes are triggered by asset sales that happen in year *t*, there is a planning deadline for capital asset sales at the end of the year. Therefore, information gathering

[12] This review of the rules regarding capital gains taxation is by necessity simplified and leaves out some details regarding capital gains taxation. For example, tax issues such as the taxation of carried interest, depreciation recapture rules, special rules and rates for collectible and personal-use capital assets and qualified small business stock all add complexities about which taxpayers or their agents will have to acquire information to be able to both optimize behavior and comply with the law.

[13] For most of the period we study, the top rate was 15%, but this changed in 2013 due to ATRA.

[14] The strategic timing of one's death can also mitigate the estate tax (Kopczuk and Slemrod, 2003).

in order to achieve the best after-tax outcome with regards to capital asset sales will have to happen by December 31 of year *t*.[15] When capital gains tax rates increase between year *t* and year *t*+1, December 31 is also the last date to realize capital gains at the year-t tax rate. As such, information gathering in order to shift capital income through time in response to a rate change will also need to happen by December 31st of the end of the year.

3.2 The Decision to Buy, Sell, or Hold, and How Taxes Affect It

A rational investor should maximize after-tax utility, which would entail taking capital gains taxes into account in deciding whether to buy, sell, or hold a capital asset. First consider the purchase decision in isolation from the rest of an investor's portfolio. The taxation of the capital gains tax structure reduces the expected after-tax rate of return to a capital investment, where given the rules discussed in Section 3.1, the reduction depends on the expected appreciation, the expected holding period, and the likelihood that an asset with appreciation can be held until death. Because higher capital gains taxes reduce the attractiveness of assets expected to appreciate in value, one may expect that the level of asset prices would react negatively to unexpected news about tax increases (described as the capitalization effect). Of course, the effect of acquiring information about capital gains taxes depends on how it changes prior beliefs. One can imagine a potential investor being pleasantly surprised to learn about the preferential lower tax rate and step-up basis at death, or being discouraged upon learning that any tax at all applies upon the sale of appreciated assets.

The decision of if, and when, to sell a capital asset should also be affected by the tax system. Certain provisions in the tax code (for example, a lower tax rate for long term capital

[15] There are some exceptions. For example, if a taxpayer sells shares at a loss and, within a 60-day period after that sale, purchases substantially similar securities, the sale is treated as a "wash sale" and the unrecognized loss is added to the basis of the purchased shares. This removes the benefit from having accelerated the losses from selling the shares. This is a case in which action taken in year *t*+1 may affect the tax liability of the taxpayer in year *t*.

gains), may encourage taxpayers to postpone an asset sale in order to obtain the favorable tax treatment. Likewise, the annual nature of tax compliance may also create annual rebalancing of portfolios to achieve a favorable mix of capital losses and gains (such that the losses almost exactly cancel the gains), and may encourage asset sales.

3.3 Evidence

The empirical evidence is largely consistent with taxpayers altering both the nature and the timing of transactions in order to achieve the greatest possible after-tax return on their capital investment. For example, as capital losses must generally be offset with capital gains, large downward movements in the stock market often leave investors with large unrealized capital losses that must be carried forward until years when these investors realize capital gains. Individuals can carry capital losses forwards indefinitely, without interest. This policy leads to the well-known strategy for minimizing capital gains tax liability of “loss harvesting,” selling capital assets with built-in losses to offset the gains of assets you have sold in the year. For example, Ivković, Poterba and Weisbenner (2005) find evidence of tax-loss selling, especially at the end of calendar years. Poterba and Weisbenner (2001) find that this is especially prevalent in years when changes in tax policy provided additional incentive to harvest losses at year-end. While tax-loss selling occurs in the corporate equity market, there is also evidence of such activity in municipal bond closed-end funds (Starks et al, 2006) and the market for long-term government and corporate bonds (Chang and Pinegar, 1986).

The preferential treatment of long-term versus short-term capital losses also appears to change investor timing of asset sales. For example, Hanlon and Pinder (2013) find that 12 months after a firm holds its initial public offering, there is an inordinate amount of sellerinitiated trading, consistent with investors waiting the full 12 months before selling their shares in order to qualify for long term capital gain treatment.

While tax-loss harvesting may cause taxpayers to sell assets they otherwise would not have sold, the fact that capital gains are only taxed upon realization also causes some taxpayers to delay recognition of a taxable gain by delaying the sale of an asset with built-in capital gains, leading to a “lock-in” effect.[16] An old and large body of literature documents this “lock-in” effect (see, for example, Auerbach (1988), and Burman and Randolph (1994)) and suggests that the presence of a capital gains tax causes individual investors to delay the sale of appreciated assets because a sale would trigger a taxable event. Ivković, Poterba and Weisbenner (2005) find that individual investors held assets with capital gains longer when trading in taxable accounts relative to their tax-deferred accounts. Further, there is evidence of taxpayers delaying the sale or gifting of appreciated capital assets until their death (Poterba, 2001).

More recently, Dai, Maydew and Shackelford (2008) also find evidence of the lock-in effect. They examine market returns in the week after an unexpected decrease in the capital gains tax rate resulting from the Taxpayer Relief Act of 1997. They find that in the week after the tax rate decrease, firms with larger built-in capital gains and a higher percentage of taxable shareholders have lower market returns, relative to other stocks. This finding is consistent with non-tax-exempt investors with large built-in gains taking advantage of the lower tax rate just following the rate decrease and selling shares, putting downward pressure on the stock price, and resulting in lower returns in the week following the capital gains tax rate decrease.

4 Data

[16] In this case, the capital gains tax has the opposite effect as in the case of loss harvesting, wherein investors sell assets they would have otherwise not have sold as a result of the tax (these two effects are described in Stiglitz (1981). The lock-in effect also has exactly opposite the effect to that of the disposition effect discussed in Odean (1998).

Because information-seeking happens through many channels, we examine a variety of measures of information search, from both IRS administrative and publicly available sources.

4.1 Google Trends

We first examine data on tax-related information search from Google Trends. Using query data on Google searches, Google Trends provides a measure of the "propensity to search" for a given search query or set of queries. More specifically, an observation in the Google Trends data will be, for a given day and geographical region, the number of Google searches for the specified search terms divided by the total number of Google searches on any topic in the time and place.[17] For our purposes, the search terms will be a broad set of capital-gains or capital-gains-tax-related search terms, and the geographic region is the United States. The qualitative results we discuss are robust to other similar sets of search terms, including simply the search term "capital gains tax."[18] After obtaining the propensity to search, Google scales the variable from 1 to 100, where the number 100 corresponds to the day with the highest search volume for this set of search terms in the entire sample period. Our sample period consists of January 1, 2004 (when Google Trends data becomes available) through March 30, 2013.

An advantage of the Google Trends measure of information search is that it is very broad. One uses Google to find information from a variety of sources, including the IRS, Wikipedia, and other online repositories of information. For example, in our sample period, Google is most often the top referring sight for visitors to IRS.gov, sometimes with as many as double the number of referrals as the second most referring site (which is often another search engine, such as Yahoo). Table 7 contains the top referring sites for all IRS.gov URLs, with the number of referrals from November, 2012 to April, 2012 (the last six months of available data).

---

[17] For more detail about Google Trends, see http://support.google.com/trends/?hl=en and the appendix to this paper.
[18] For narrower and/or more obscure search terms, the data will be missing for some days, because Google Trends does not provide data for search terms with relatively low volume.

A disadvantage of the Google Trends measure is that changes in the measure can be difficult to interpret precisely. We do not know how many more searches correspond to an absolute increase of the measure by, say, 10.[19] It is also possible that changes in the measure over longer time frames are driven by changes in total Google usage (the denominator of the "propensity" variable) over time rather than decreases in the number of queries (the numerator). This last concern is alleviated by focusing primarily on high-frequency variation, as overall Google usage is in most periods not likely to swing rapidly from day to day.[20,21]

4.2 Wikipedia Page Views

Our second source of publicly available data related to information search is page view data from Wikipedia, a free online encyclopedia edited by Internet users. Using Wikipedia data follows naturally from using Google Trends data, as entering a given term into Google will frequently yield a Wikipedia page related to that term as one of the first (and often the first) suggested sites to visit. We have obtained data, by hour, on page views to the English language version Wikipedia site "Capital Gains Taxation in the United States". A page view consists of any time a browser requests to load the web page in question.[22] We aggregate this data up to the

[19] Eliminating the second caveat by assuming the denominator is unchanging, we can interpret an increase of 10 in the measure as an increase in the number of queries of 10 percent of the peak number of total searches in a day in the sample period.

[20] A possible exception is systematic day-of-the-week seasonality in Google usage, e.g. if people search more on Mondays than Sundays. Throughout our analysis, we account for these effects and weekly variation in search volume via day-of-the-week fixed-effects.

[21] Notably, the Wikipedia data is *not* scaled in the same way. We have replicated the results in Sections 6 and 7 using a measure of Wikipedia usage scaled in a similar manner to how Google data are scaled. To do this, we use page views of the Wikipedia main page as a measure of general Wikipedia use. This replication further assures us that variation in total Google search from day to day will not confound the results.

[22] Visits, in contrast, are a web analytic that count the number of page views from unique Internet Protocol (IP) addresses. One weakness of the Wikipedia view count data is that Wikipedia does not track visitors. So, to the extent that one viewer views the same webpage many times in one day, this will alter our results. We see no reason why one viewer would return to a tax-related Wikipedia page enough times in a day to materially affect our data. As we show in Section 6, the notable spikes in views of this Wikipedia page also occur at the same time as spikes in Google Searches, which alleviates this concern. We have also examined page views of the page "Capital Gains Taxation," and note that it is highly correlated with views of "Capital Gains Tax in the United States."

daily level (summing the 24 hours in a day) to obtain our measure of Wikipedia page views.[23] We use Wikipedia data from January 1, 2008 to March 30, 2013. One disadvantage to the Wikipedia data is that it is available for a substantially shorter time period than is the Google data. However, on the upside, we are able to obtain the raw number of views of the webpage, and so the interpretation of the size of changes in information search is simplified.

4.3 IRS Phone Calls

Our third source of data regards aggregate calls made to the IRS's toll free phone number, where taxpayers can call and speak to a representative from the IRS or listen to prerecorded messages. In the course of a phone call with the IRS, some taxpayers will listen to prerecorded messages about various "tax topics." Examples are as varied as tax topic 304, "Extensions of Time to File Your Tax Return", tax topic 151, "Your Appeal Rights", or even tax topic 417 "Earnings for Clergy." The IRS phone call data we analyze consists of the number of phone calls, daily, that access the IRS tax topic 409, "Capital Gains and Losses,"[24] and the total number of calls to the general inquiries phone number. Taxpayers may reach the tax topics function in a number of ways, including calling the tax-topics-specific phone number, or being referred to the topic after a conversation with a representative or an interaction with the automated system. We use data from February 1, 2002 to March 31, 2012 (the time series available from the IRS).

4.4 IRS Website Visits

The last data set we analyze also comes from the IRS, and consists of the aggregate daily number of visitors to the IRS's webpage. The IRS maintains a website, IRS.gov, which has a

[23] For a very limited number of hours (or days) days in the time series, Wikipedia data is not available. This is due to resource allocation within Wikipedia itself—in rare occasions when Wikipedia is conducting activity or making site changes that require nearly all of its computing resources, low priority tasks such as tabulating historical statistics are simply not completed. In cases where the entire day is omitted, which occurs for just 15 days, we use the previous day's value. In cases where data on certain hours during a day is unavailable, which occurs for 128 days, we scale the observation up by 24 divided by the number of hours that are available.

[24] The text for this prerecorded message can be found at http://www.irs.gov/taxtopics/tc409.html

vast amount of general tax information, reports, forms and press releases from the IRS, and other material. Like the Wikipedia data, this coincides well with our use of Google search data. For example, searching for "capital gains taxes" on Google yields, in the first five search results provided, the web pages from both Wikipedia and IRS.gov. Google is frequently the top referring website to IRS.gov, and Wikipedia is often in the top 60 referring web pages. We obtain, directly from the IRS, the number of page views and visits, by day, to any site hosted by IRS.gov. We are able to obtain this data for the entire time series for which the IRS has maintained the data, February 1, 2002 to March 31, 2012. Analysis of IRS.gov traffic carries with it the benefit of analyzing taxpayer information search about a very broad set of tax information. However, this measure of information search is not limited to information search about capital gains taxes specifically.[25]

5 Estimation Procedure

Let I_{it} denote information search on date t from some source of information i, such as Google, Wikipedia, or the IRS call line. We wish to estimate the effect of several different events on information search through source i. The events we study may be either non-recurring, as in the case of a Presidential debate that mentions capital gains taxes, or they occur annually, as in the case of tax return filing season. Non-recurring events are indexed by the subscript k, and annual events are denoted according to the day of the year on which they occur, denoted *DoY*.

As we discuss in more detail later, the data on inquiries display marked seasonality at the yearly and weekly levels. In order to evaluate whether a particular (non-seasonal) event increased information search on date t, we must properly specify the counterfactual level of search, conditional on the day of the week and day of the year of date t. For example, suppose we want to know the impact of an event occurring on March 1st, 2010. Suppose further, as we show

[25] We are hopeful of obtaining high-frequency, topic-specific data about the IRS website traffic in the future.

later, that searches are elevated in early March, compared to other times of year (because March occurs during tax filing season). If we simply regressed information search on a dummy variable for March 1st, 2010, we would then tend to over-estimate the impact of this event if we did not control for yearly seasonality. We control for the day of the week for the same reason: all of our measures of inquiries are higher on weekdays than on weekends, so a consistent estimate of the impact of an event should control for whether that event occurred, for example, on a Monday or a Saturday.

Unlike for much analysis of time-series data, yearly seasonality is of intrinsic interest for our research questions. In particular, increases in information-seeking during tax season will affect the form of the yearly seasonality. Information searches occurring at the end of the year can be thought of as response to a time notch. To address this concern, we employ a method that both estimates the seasonal patterns and allows us to perform classical statistical tests for whether information search is significantly higher on a particular day of the year than the average level of information search. Specifically, for each information search series *i* we estimate the function:

$$I_{it} = \sum_k \left[\beta_{k0}F_{kt} + \beta_{k1}F_{k(t-1)} + \cdots + \beta_{k4}F_{k(t-4)}\right] + x_t'\gamma + f_i(DoY_t) + \delta_{DoW,i} + u_{it} \qquad (1)$$

The term inside square brackets is a set of dummy variables: $F_{k(t-j)}$ equals 1 if event k occurred on date *(t-j)* and zero otherwise. The four-day event window was selected by noting that when large spikes in information occur in the data, search levels return to baseline within four days. We also include a vector of continuous, time-varying linear covariates x_t'—such as trading volume on the stock market[26]—along with a general non-linear function in day-of-the-year $f_i(DoY_t)$,[27] and a day-of-the-week fixed effect, $\delta_{DoW,i}$.

[26] Stock market data are not generated when the market is closed on weekends and holidays. When we run

One natural alternative way to estimate the function $f_i(DoY_t)$ would be to include day-of-the-year fixed effects. However, with only 5 to 9 years of data for each type of inquiry, day-of-the-year fixed effects would be imprecisely estimated. Intuitively, though, we should be able to use data on $f_i(DoY_{t-1})$ and $f_i(DoY_{t+1})$ (the day before and the day after a specific day of the year) to estimate $f_i(DoY_t)$, under the assumption that $f_i(DoY_t)$ should not change *too* sharply from day to day.

We proceed by employing a kernel-weighted local linear regression. The strength of this estimator is that 1) it estimates a smooth trend in day-of-the-year, which allows us to control for day-of-the-year when estimating the effect of events, and 2) it increases the precision of the estimates of $f_i(DoY_t)$ relative to fixed effects. The potential downside to the estimator is that the assumption that $f_i(DoY_t)$ does not on average change *too* sharply from day to day may be wrong (especially, for example, around April 15th). The strength of this assumption is governed by the bandwidth of the estimator, and also by our choice of kernel density function. We discuss these choices below. Our preferred choices for kernel density function and bandwidth produce estimates of seasonal trends that match the pattern implied by fixed-effects estimates.

In order to consistently estimate equation (1), which contains a non-linear function and a set of linear covariates, we use the double residual regression method suggested by Robinson (1988), also discussed in Hardle and Linton (1994). This first step of this three-step estimator consists of several non-parametric regressions of the following form:

$$E[X_t] = f_i(DoY_t),$$

where X_t is the dependent variable or one of the linear covariates, i.e. I_{it}, F_{kt}, x_t or $\delta_{DoW,t}$. We then obtain the residuals from this regression, which we denote with star superscripts. The residuals

regressions that include trading volume data (Sections 8.1 and 8.2), we omit weekends and holidays from the analysis (and thus omit two day-of-the-week fixed effects and holiday dummy variables from the specification).

[27] For simplicity, leap days are dropped from the sample so that *DoY* takes uniform values every year.

represent the component of I_{it}, F_{kt}, x_t or $\delta_{DoW,t}$ not correlated with the general within-year pattern represented by $f_i(DoY_t)$. The second step estimates the linear components of the model consistently using ordinary-least-squares regression on these residuals:

$$E[I^*{}_{it}] = \sum_k \left[\beta_{k0}F^*_{kt} + \beta_{k1}F^*_{k(t-1)} + \cdots + \beta_{k4}F^*_{k(t-4)}\right] + x_t'^*\gamma + \delta^*_{DoW,i} \quad (2)$$

The third step regresses the residuals from the estimation of equation 2—denoted by $I^{**}{}_{it}$ — non-parametrically on DoY_t to obtain a consistent estimate of the function $f_i(DoY_t)$.

$$E[I^{**}{}_{it}] = f_i(DoY_t) \quad (3)$$

The first and third steps involve non-parametric regression, which Robinson (1988) suggests implementing via kernel-weighted local polynomial regression. In this paper, we use local linear regressions with a Gaussian kernel density function and a bandwidth of four days.[28] This procedure estimates a weighted OLS regression of the dependent variable on *DoY* at each value of *DoY* in the data, where the weights are determined by the kernel density function and the bandwidth. The bandwidth was selected to visually match the fixed-effects estimator of the function $f_i(DoY_t)$, but a data-driven choice of bandwidth—specifically selecting the bandwidth that minimizes the conditional weighted mean integrated squared error—yields nearly identical results. The estimates we present are also virtually unchanged by varying the degree of the local polynomial, the bandwidth, and/or our choice of kernel function, with the exception that a wider

[28] This bandwidth and kernel density function applies to non-parametric estimations from the third stage of the procedure, and nonparametric estimations involving continuous variables in the first stage. For discrete variables in the first stage, such as event dummy variables, we use a bandwidth of zero, to reflect that there should be no smoothing at this stage. Another complication is that traditional statistical packages and programs will implement the smoothing on a linear variable, while day-of-the-year is a cyclical variable. That is, January 1 and December 31 should be adjacent for smoothing purposes. Ignoring this problem results in a discontinuity in the seasonal pattern between these two days. We eliminate the discontinuity by estimating the seasonal pattern twice: once where the discontinuity is imposed at January 1, and a second time where the discontinuity is imposed at the 200th day of the year (July 19th). Then, we replace the 10 days around January 1 from the first estimation with these days from the second estimation.

bandwidth results in a smoother function that no longer resembles the fixed-effects estimates and a narrower bandwidth results in a more jagged function.

For a recurring annual event such as a filing deadline, we can examine the function $f_i(DoY_t)$ on days-of-year corresponding to the annual event to understand the effect of the event on information search. The comparison that seems most natural is to test the hypothesis that, on a particular day of the year, DoY', inquiries are higher than their unconditional average, i.e. that $f_i(DoY') > E(f_i(DoY))$.[29]

Our estimate of the function $f_i(DoY)$ will inform us about the importance of tax season for information search,[30] but it may not capture all variation due to compliance deadlines. For example, the mid-April filing deadline does not occur on April 15 of a given year if that date falls on a weekend or Emancipation Day. Instead, it can occur as late as April 18 in some years. We account for variation in compliance dates by year by including dummy variables for compliance events in addition to the locally linear function. Our estimate of the function $f_i(DoY)$ will estimate the average effect on April 15, but we add an additional dummy variable for the precise date of the filing deadline to capture variation that occurs specifically on the deadline each year. Doing so does not significantly change our picture of the importance of tax season for inquiries, but it does highlight the sharp spike that occurs exactly on the mid-April deadline each year. The inclusion of this dummy variable also helps mitigate the smoothness assumption in estimating $f_i(DoY)$, effectively imposing a bandwidth of zero on the date of the filing deadline. To further reflect that the smoothness assumptions should not apply when there are sharp

[29]This is the same hypothesis test we would carry out if we were including day-of-the-week fixed effects, and inquiries on the “left-out” DoY was a date on which searches were on average equal to the average over all days.

[30] For Google data, the interpretation of the estimates is somewhat complicated by the fact that the dependent variable is a measure of propensities to search. If there were yearly seasonality in overall Google searches, this would affect our estimate of $f_i(DoY)$ by decreasing it on days of the year when Google usage was highest. The fact that the pattern estimated by Google data and other measures of information search are similar largely alleviates this concern. The inclusion of day-of-the-week fixed effects also makes our results robust to weekly seasonality in Google searches.

changes on particular days of the year, we also include dummy variables for the following holidays: New Year's Day, President's Day, Martin Luther King, Jr. Day, Memorial Day, Independence Day, Labor Day, Thanksgiving, and Christmas.

6 Exploring the Raw Data

Figure 1 plots the evolution of three measures of information search on capital gains taxes over time. Calls to the IRS hotline inquiring about capital gains taxes occur almost exclusively during tax season, from mid-January to mid-April. The absolute volume of calls occurring during a tax season diminishes considerably throughout the time period, from 2002 to 2012. Wikipedia page views, in contrast, increase over time. These patterns are consistent with online information having supplanted information obtained by telephone as Internet access increased markedly over the sample period. The Google Trends measure displays a slight downward trend. Recall that this does not mean that Google searches of capital gains taxes decreased over time, only that the *share* of Google searches that concerned capital gains taxes decreased over time. Visually apparent in each time series is a strong pattern of yearly seasonality. This is most pronounced in the IRS call log series, where the yearly seasonality drives virtually all of the variation in the series over time, but the pattern is also present in Google searches and Wikipedia page views. Zooming in on a narrower time frame also reveals strong weekly seasonality in each time series.

To focus on high-frequency variation, for each information-search series we detrend the data using a Hodrick-Prescott filter.[31] Next, we seasonally adjust the data, removing weekly seasonality by regression on day-of-the-week fixed effects, and removing yearly seasonality using the smoothed-fixed-effects-with-holidays method described in section 5.[32] Figure 2 plots

[31] Whenever we use a Hodrick-Prescott filter in this paper, we use a smoothing parameter of 10^7. This value was selected by trial and error, with the goal that the long-run trend capture long-run movements in the series but not variation due to yearly seasonality.
[32] The estimate of seasonality we use to examine yearly seasonality in the data comes from the regressions run in the

detrended and seasonally adjusted time series for the entire 10-year period. Seasonal adjustment removes most of the variation in the IRS call log data.[33] The Google and Wikipedia series, however, contain marked spikes around late October of 2008, January 24, 2012 and January 2, 2013. We attribute the first of these to a combination of the stock market crash and the presidential election in 2008, the second to the release of presidential candidate Mitt Romney's tax return on January 24, 2012, and the last to the passage of the American Taxpayer Relief Act, which resolved what was commonly called the "fiscal cliff" debate and enacted an increase from 15 to 20 percent of the top rate on long-term capital gains.

To allow for visual comparison of the three measures of information search, we normalize each variable by dividing by its standard deviation.[34] Figure 3 plots the estimated yearly seasonality in the standardized data. For each measure of information search, we observe a sizable and significant increase in search behavior during the period commonly known as tax season, and an even stronger spike in the immediate run-up to the April 15 filing deadline. Right after April 15, information search drops off sharply. Clearly, the desire to comply with the tax law as one fills out a tax return leads taxpayers to search for information. The specific pattern we observe—that searches are heightened during tax season, but rise sharply in the few days before the deadline—also matches the pattern in the timing of filing documented by Slemrod et al (1997).

In the next set of figures, we focus more narrowly on variation in the standardized, detrended, and seasonally-adjusted series around the three dates where one observes large spikes

---

next section, and controls for the impact of different events that caused large outliers to occur in the series. With a longer time series or less extreme outliers, controlling for outliers would not be necessary, but with a short time series, a one-time spike in the data can bias the seasonal pattern.

[33] The volatility of the call log time series is also seasonal, so there remains some variation in the time series during tax season that is not completely removed by the methods we use to control for yearly seasonality.

[34] Because the volatility of the IRS call log series varies significantly over time, we standardize it by dividing by its standard deviation *by year.*

in information search. Figure 4 focuses on the spike in October and early November 2008. There are two obvious candidates for elevated information search in this time period. The first of these is the presidential election of 2008, and the second is the financial crisis and accompanying stock market crash of 2008. Because capital gains taxes were an issue on which candidates John McCain and Barack Obama differed substantially, the election could influence search behavior due to either the desire to make an informed choice about asset purchases or sales based on expected future tax policy, or due to the desire to understand the consequences for future tax policy of a President McCain or Obama.

The market crash of 2008 resulted in large capital losses for many investors and, due to the extreme volatility of the market, potentially also large short-run capital gains. Our tentative conclusion is that both the market crash and the election played some role, although it is difficult to disentangle the effect of the two events. Our conclusion is based on the fact that the largest swings in the stock market in this period (marked by red and black vertical lines in figure 4) are associated with spikes in search behavior,[35] and that information search surged around the date of the presidential election itself. We present other evidence on how the stock market crash might have affected information search by focusing more narrowly on capital losses in section 8, and Figure 6 provides further evidence on presidential elections.

Figure 5 plots the data from Figure 2 around the spike in information search observed in January, 2012. We attribute this surge in information search to the release of presidential candidate Mitt Romney's 2011 tax return on January 24.[36] The tax return generated substantial

[35] One of these dates, October 15, 2008, was the same date as a presidential debate in which both candidates discussed their proposals for capital gains taxes. This further limits our ability to draw inference about what caused elevated information search in this period.

[36] An earlier spike occurred after Mr. Romney announced that his effective tax rate was around 15 percent, on January 17, 2012. This announcement also generated substantial news coverage.

news coverage, in part due to his low effective tax rate.[37] As most news articles on the subject noted, much of Romney's income came from the realization of long-run capital gains taxes, which were subject to a maximum personal tax rate of 15 percent. Another possible explanation for the spike in search behavior on this date is the State of the Union Address, which also occurred on January 24, 2012. In his speech, President Obama advocated for taxing the wealthy at higher tax rates (supporting the "Buffett Rule"), but he mentioned neither capital gains taxes specifically nor investment taxes more generally.[38] It is, however, possible that Obama's rhetoric on taxing the rich led the public to pay greater attention to the release of Mitt Romney's tax return, and thereby amplified its effect on capital-gains-tax-related information search.[39]

Figure 6 plots the data from Figure 2 at the end of 2012 and the beginning of 2013. A presidential debate between Barack Obama and Mitt Romney appears to have sparked considerable information search. As in 2008, we also see elevated search immediately following the presidential election, perhaps as voters and investors researched what might happen to capital gains taxes in the aftermath of President Obama's re-election. Finally, the largest outlier in information search in the 10-year period covered by our data—an increase of just over 4 standard deviations in Wikipedia page views and over 6 standard deviations of Google searches—occurred on January 2, 2013. We attribute this to the passage of the American Taxpayer Relief Act on that date. This bill partially resolved what was commonly called the "fiscal cliff" debate, and increased the top marginal tax rate on long-term capital gains from 15 percent to 20 percent. Our evidence strongly suggests, therefore, that individuals search for information both in

[37] See, for example, http://www.nytimes.com/2012/01/25/us/politics/romneys-tax-returns-show-21-6-million-income-in-10.html?pagewanted=all&_r=0.

[38] The "Buffett Rule" is a tax plan proposed by President Obama, wherein individuals making over $1.0 million in taxable income would be subject to a minimum average tax rate of 30%.

[39] Further insight into this can be gained by examining intra-day search activity, available only for the Wikipedia measure of search activity. Search activity started rising dramatically mid-day on the 24th, consistent with Romney's tax return being at least partially responsible for this increase (as Obama's speech was not delivered until that night).

response to policy changes and in response to potential policy changes signaled by political events.

7 Regression Analysis of the Impact of Events

We start with the detrended and seasonally *unadjusted* time series. The regression procedure we use, outlined in Section 5, explicitly controls for variation due to weekly and yearly seasonality. Table 2 describes the events we study, which are also discussed in the previous section.[40]

Table 3 describes the estimated impact of events on information search through Google, Wikipedia, and the IRS web page. To examine statistical significance, we perform F-tests on the β's from equation 1, which generates a *p*-value that corresponds to the probability that changes in information search of the magnitude we observe would have occurred at random during the event window we specify. The results suggest that the events we study each have a large and significant impact on information search through Google and Wikipedia, but not through the IRS call line.[41] When we estimate the overall impact of the event rather than examining a single-day impact, the release of Mitt Romney's tax return surpasses the passage of ATRA as the one-time event that generated the most taxpayer information search, through both Google and Wikipedia. This occurs because the effect of ATRA passage declined more rapidly, fading in two days

[40] There is modestly elevated information search via Google, Wikipedia or both, on a few dates that we do not include in the analysis. These dates include January 18, 2004 (a state of the Union address by George W. Bush advocating the extension of capital gains tax cuts), November 2, 2004 (re-election of George W. Bush), May 16, 2006 (the extension of the 2003 capital gains tax cuts), January 27, 2010 (a State of the Union address by Barack Obama advocating a cut in capital gains taxes for some taxpayers), and March 23, 2010 (passage of the Affordable Care Act, which included a "net investment income tax"). In each case, the response of searches is qualitatively similar—a spike in searches that fades in three to four days—but quantitatively smaller than the events we do include in the formal analysis.

[41] Two events are estimated to have an effect on calls to the IRS. However, seasonal variation in the volatility of the calls measure means that the statistical significance procedure is biased for dates occurring during tax season. As such, we are cautious about interpreting this episode.

instead of four.[42] Notably, ATRA passage was anticipated in the days leading up to January 2. The bill passed Congress on January 1st, and the anticipated increase in the top capital gains rate likely generated income shifting from 2013 to 2012, which may have also caused information search. From Figure 6 we can see that elevated information search occurred on the two days prior to ATRA passage, December 31 and January 1st. Adding these dates to the event window increases the estimated impact of ATRA by approximately 5 standard deviations for Google and Wikipedia, but changes little else.

8 Relating Tax Information Search to Stock Market Activity—Toward Establishing Causal Directions

A key advantage of focusing on capital gains taxation is the availability of high-frequency data on many capital asset sales, specifically sales of stock. Relating these data to data on information searches holds the promise of better understanding the causal connections between information search and capital-asset-related behavior. After all, we are not only interested in what causes people to search for information, but also in to what extent the acquisition of information affects behavior, in this case behavior related to the sale, purchase, and holding of capital assets.

8.1 The Lead-Lag Relationship to Volume, Volatility, and Market Return

Our first strategy is to examine the lead-lag relationship between measures of behavior and information searches. If searches lead behavior, then we have reason to pursue the idea that the information obtained affected subsequent decisions. We investigate two new data series, all which represent general stock market activity.[43] The market-related measure we use is trading volume. We obtain the dollar value of shares traded from all publicly listed firms from the Center

[42] Figure 6 also suggests that some of the impact of ATRA passage may have occurred before the event date, in which case we underestimate its impact.
[43] Ideally, we would use realized capital gains aggregated at a daily level, but these data are not yet available to us.

for Research in Security Prices (CRSP),[44] which we use as a measure of broad market activity. Note that these transactions include many where the buyer and/or seller is not subject to capital gains taxes, may not be a human trader capable of information search, or where the asset does not have an accrued gain that will be subject to taxation. These possibilities do not threaten our identification so long as the percentage of transactions that is not subject to capital gains taxes does not change substantially from day to day.

Having obtained daily time series on trading volume, we detrend the measure using a Hodrick-Prescott filter, and include the log of daily trading volume in the regression described in equation (1). For our baseline regression, we include five days of leads and lags of log trading volume.[45] These regressions control for the events documented in Table 2 and for weekly and yearly seasonality. Columns (1) through (3) of Table 4 report the coefficients on standardized data.

The results indicate that a 1 percent increase in trading volume on date t is associated with a 0.0029 standard deviation increase in Google searches and a 0.0031 standard deviation increase in Wikipedia Page views, on date t. These effects are significantly different from zero at the 10 percent level (p = 0.068 for Google, p = 0.055 for Wikipedia). Given that the standard deviation of log trading volume is 0.2233, a one-standard-deviation increase in trading volume should be associated with a 0.064 standard deviation increase in Google searches, and a 0.068 standard deviation increase in Wikipedia Page views. The relationship between trading volume

[44] Specifically, we sum, by day, the variable "VOL" from the dataset crsp.dsf provided by Wharton Research Database Services (WRDS).

[45] Stock market data do not exist when the market is closed for weekends or holidays. As mentioned above, we omit weekends and holidays from this part of the analysis. The leads and lags we use are market-dated leads and lags, so for example the first lag of an observation on a Monday will be the observation on the previous Friday. If we instead assign Saturdays and Sundays a value of log trading volume from the previous Friday, we obtain nearly identical point estimates, with two differences: a significant two-day lag and lead effect emerges and the p-values on coefficients are lower. The first is an anomaly from this treatment of weekends, and the second a byproduct of increased sample size.

and information search suggests that when trading volume is high, individuals seek information about capital gains taxes. We find that searches on date *t* are not significantly related to trading volume on date *t*+1 or beyond. All of the effect of market movement on information search occurs on the same date as the market movement, or one day before. As a result, we are not able to determine from these data whether individuals seek information on capital gains taxes primarily before or after they make a decision regarding the sale of a capital asset. If they seek information before making a decision, they do so less than a day in advance, as far as these data are able to tell us.

Much of the variation in these results may be driven by the 2008 stock market crash, a period of extremely high volatility and high trading volume, and widely varying returns. Columns (4) and (5) of Table 4 provide the estimates of the same regression specification, but limiting the inclusion of market variables to the period from September 2008 to February 2009. Columns (6) and (7) provide the estimates of the same regression specification, instead excluding this extraordinary period. For both Google and Wikipedia data, the estimated effect during the extraordinary period is much larger, although imprecisely estimated. When excluding the extraordinary period the estimate is slightly smaller in Google data and slightly larger in the Wikipedia data. Overall, these results suggest that search activity may have been particularly sensitive to changes in trading volume during the stock market crash (which is consistent with the findings in Figure 4), and that the results are not exclusively driven by a single period during which the stock market behaved very atypically.[46]

[46] The magnitude and significance of the results are also not altered substantively by 1) the inclusion of lags of the dependent variable, or 2) the inclusion of daily measures of the absolute value of stock market returns (VWRETD in CRSP), the outright value of stock market returns, and a measure of volatility perceived by futures markets (VIX in CRSP). One puzzling result is that for Wikipedia Page Views, we sometimes see negative and significant coefficients on either VIX or VWRETD when these are included in the regression along with trading volume. We suspect these are driven by strong correlations between these variables and trading volume, but have been unable to fully explain their statistical significance.

8.2 Do taxpayers search for capital gains tax information when they are contemplating buying or selling stock?

In the previous section we explored stock volume as an indicator of, in the language of Zerbinos (1990), "when a person recognizes a gap in their knowledge that may motivate that person to acquire new information." Volume is presumably associated with the extent to which taxpayers are contemplating buying or selling an asset.

In this section we pursue an alternative indicator of the taxpayer demand for capital gains tax information that is itself based on observed search volume. In particular, we use Google Trends to obtain a measure of the volume of searches for phrases related to personal investment advice such as "stock advice", "should buy stock," "should sell stock," and "investment advice."[47] Note that, although Google searches for investment advice can lead one to websites that purport to provide such advice, searches for "capital gains tax" generally do not.

Table 5 shows the results of including the investment advice measure as a variable explaining search volume for capital gains tax information, both in addition to stock trading volume and as a replacement for stock trading volume. Several results of interest emerge. First, consider the results for Google searches. When investment advice search volume is included as an alternative to stock trading volume, the same positive contemporaneous association appears, and the effect of a one standard deviation change on capital gains tax related searches is 40 percent lower. In a "horse race" when both stock trading volume and investment advice search volume are included as explanatory variables, unsurprisingly the statistical significance of the former declines because of the high correlation between the two variables. In the horse race, investment advice search volume wins, retaining its significance while stock trading volume losses its own significance. Strikingly, the one-day lead and lag values of stock investment

[47] The full list of search terms is in the appendix to this paper.

advice search volume are significantly positively associated with capital gains tax search volume; recall that this did not occur for stock trading volume as a measure of the demand of tax information. This estimated association is consistent with a story that taxpayers first recognize the need for information regarding buying or selling stock and, in the process, learn that relevant to this decision are the tax consequences.

The key advantage of the stock investment advice search volume variable is that it arguably captures the extent of taxpayer demand for information for which capital gains tax knowledge is crucial. One potential disadvantage is that in regression analysis one may pick up any shocks that affect all Google Searches. To that point note in Table 5 that the same pattern of results also applies when the volume of Wikipedia searches is the dependent variable, with one exception: the lead relationship highlighted above, although positive, does not reach statistical significance.

8.3 Losses

Because the capital gains tax rules related to the sales of assets with capital losses are especially important for many tax minimization strategies, such as loss harvesting, we also construct two measures of information search related specifically to capital losses. The first is weekly Google searches for the phrase "capital loss," and the second is monthly searches related to capital losses on the IRS website, IRS.gov. We obtain the latter measure directly from the IRS. On IRS.gov, there is a search functionality, where website users can search for items they would like to know more about. The IRS provided to us, at a monthly level, the most frequent search terms, and the number of searches for all terms with the word "capital", and "loss", in the search.[48] These time series are plotted in Figure 7.

[48] The specific search terms people used were "capital losses," "capital loss," "capital loss carryover," "capital gains and losses," and "capital loss carryover worksheet."

As with searches for capital gains, much of the variation in both measures of information search is seasonal: people tend to search for information regarding capital losses during tax season. There is also typically a small increase in searches in December of each year, when some taxpayers harvest capital losses to reduce their tax liability. Interestingly, searches for information on capital losses increase dramatically during October 2008, and surge even further in December of 2008. We likely observe this because the 2008 stock market crash caused taxpayers to search for information on capital losses. When the crash began in October, investors began to research the tax implications of the unrealized or realized losses they had sustained, perhaps evaluating the merits of pulling their wealth out of (or investing in) the declining stock market. Of particular interest is that, for an investor who had lost money in the crash, harvesting capital losses before the end of 2008 could reduce the taxpayer's tax liability substantially (if the investor also had capital gains to offset). Consequently, investors apparently sought information on capital losses while deciding whether to harvest losses at the end of 2008.[49]

For both Google and IRS searches, we can also see that searches for information on capital losses *during tax season* are higher after the 2008 crash than before, for three years after the crash. Undoubtedly, for several years after the crash investors realized capital losses with greater frequency than before the crash. As one might expect, information search for tax compliance purposes also increased after 2008. This constitutes additional evidence of spillovers between macroeconomic changes and information search for the purpose of tax compliance, discussed further in Section 8.4.

8.4 Delayed effects

[49] Our query of Google Trends for searches of "Capital Loss" also pointed to several news articles advising investors to realize capital losses to dampen the strain on their finances caused by the crash, such as the following article: http://www.sfgate.com/business/article/Last-chance-to-get-tax-break-with-capital-loss-3256362.php.

Not all information searches will happen immediately before or following some event. In particular, behavior and market movement during the year can be associated with information search at year end, as individuals decide whether to realize capital gains or losses prior to a time notch. To examine this possibility, we include an interaction of a year-end dummy—equal to one in the last five days of the year and zero otherwise—with aggregated yearly trading volume (in Table 6, *End of Year X Log Vol. This Yr.*) and the buy-and-hold market return for the year that is ending (*End of Year X Mkt. Ret. This Yr.*). We obtain the buy-and-hold market return using the CRSP daily return (variable VWRETD), cumulated over the entire year.[50] The yearly buy-and-hold return is a proxy for the overall market performance that year, although it does not measure exactly accrued net capital gains from the sale of stock. This allows us to examine whether individuals search more as the time notch approaches in years when it would be more advantageous to realize gains or losses in that year.

Individuals may also increase search activity during tax season of the subsequent year, as they research the compliance consequences of their previous actions. We explore this possibility by including an interaction of a tax season dummy—equal to 1 between January 30 and April 15—with aggregated yearly trading volume (*Tax Season x Log Vol. Prvs. Yr.*) and stock market returns from the previous year (*Tax Season x Mkt. Ret. Prvs. Yr.*). This allows us to examine if high trading volume and/or returns in a year are associated with greater information search for compliance purposes in the following year. The estimation controls for weekly and yearly seasonality, and the impact of events (as in Table 3).[51] We also include the yearly variables with which we interact end-of-year and tax-season dummies.

[50] To aggregate volume, we simply sum the volume by year. To aggregate daily returns R provided by CRSP across an entire year we use the following formula:
$R_{year}=\exp\{\sum_{\{t\,\varepsilon\,year\}}[\log(R_t+1)]\}-1$.

[51] The results are unaffected by the inclusion of daily variation in stock market trading volume as in Table 4. We do

The results of this estimation are provided in Table 6. High trading volume in year *y* is associated with elevated information search through Google and Wikipedia at the end of year *y*, and with elevated information search through Google, Wikipedia, and the IRS telephone hotline during tax season of year *y*+1. High stock market returns in year *y* are associated with elevated information search through Google and Wikipedia at the end of year *y*, but not with information search during filing season. This pattern is consistent with market returns affecting the gain to making tax-efficient decisions at year end—because timing matters more for realizing large capital gains than small ones—while high trading volume affects both the gain to making tax efficient decisions at year end *and* the number of taxpayers who will need to understand capital gains to complete their tax return come tax season.

9 Lessons from Aggregate Data on IRS Website and Call Line Usage

In this section we explore patterns in time series data on the use of the IRS website and call line usage which are not specific to capital gains taxes. Specifically, we use daily visits to `IRS.gov` and the number of daily calls to the IRS' Individual Assistance Telephone Line.[52] In order for their effects to be visible in broader measures of information search, events must significantly change the importance of understanding taxes at a given time, for a large number of taxpayers.

Figure 8 plots daily visits to *all* web pages in the IRS.gov domain.[53] The data are available from February 1, 2002 to March 31, 2012, with a gap in late 2003 and early 2004. Figure 9 plots daily calls to the individual tax phone line, available from September 17, 1999 to

---

not include this variable in our preferred specification because doing so results in a loss of precision from the exclusion of weekends. This loss of precision would not, however, affect the statistical significance of the results according to the 1, 5, and 10 percent benchmarks as shown in Table 6.

[52] The phone number for the line is 1-800-829-1040. This number is provided in several places on the IRS web page and in paper literature provided by the IRS such as the instructions for Form 1040, the individual income tax return.

[53] High-frequency data on visits to the IRS's web pages specifically explaining capital gains tax rules are not yet available.

July 11, 2013. Unsurprisingly, the IRS.gov domain and call line experience elevated traffic during tax season. Note also that the amount of traffic experienced during tax season increases over time for the web page, probably due to increased use of the Internet by taxpayers. The call line, in contrast, decreased in usage from 1999 to 2008, and then experienced a resurgence. Together these results suggest that the Internet has not completely crowded out the use of the call line.

Most interestingly, we observe an abnormal surge in visits to IRS web pages and calls line traffic in May of 2008. Search volume usually drops sharply after the mid-April tax deadline, but in 2008 it remained high throughout the month of May. Among the top five searches on the IRS web page during this month were "stimulus," "rebate," and "stimulus check," and many of the top pages viewed also dealt with the stimulus rebates. The tax rebates enacted by the Economic Stimulus Act of 2008 led millions of taxpayers[54] to visit the IRS web page to figure out how the federal stimulus program affected them. This corroborates the evidence provided by Sahm, Shapiro, and Slemrod (2012) that these rebate checks were a relatively salient form of economic stimulus with a relatively high (compared to a reduction in employer withholding rates) marginal propensity to consume.[55] Using the intuition of rational attention, it makes more sense for individuals to seek information in May about their stimulus check if they intend to spend it compared to if they intend to deposit it in their savings account.

10 Conclusions

[54] The precise number of taxpayers is difficult to determine, since some taxpayers may be counted as more than one visit or call. Even with conservative guesses for how many of the unique calls and website visits in May of 2008 corresponded to unique taxpayers, the number will be in the millions. For example, the spikes in searches per day during the period peaked at over 400 thousand unique calls and over 4 million website visits *in a single day*.

[55] Of course, we do not observe the role of information for stimulus funds included in monthly paychecks, so we cannot comment on the central question of Sahm, Shapiro, and Slemrod (2012), which is whether mailed rebate checks were more effective in increasing consumption than stimulus funds added to monthly paychecks.

It is well-established that in general taxpayers know very little about the US income tax, and have systematic misperceptions. Given that acquiring information is costly, it may be optimal for individuals to learn only *if* the expected return is high enough and only *when* the information is most useful, known as rational attention. Because people are learning—and forgetting—things all the time, the process of net information acquisition is critical to a dynamic understanding of tax salience. Using newly available IRS administrative data and publicly available information on Google and Wikipedia searches, this paper establishes that people seek information about the US income tax in systematic ways that are consistent with the idea of rational attention. When policies change or seem likely to change, when filing deadlines or time notches loom, people turn to online resources like Google, Wikipedia and the IRS website, as well as traditional information resources like the IRS telephone hotline, to learn how the tax code affects them. In addition, people search for tax-related information when newsworthy events make taxes more salient, and in so doing they may incidentally obtain information relevant to their own financial situation. When policy or news events generate plausibly exogenous shocks to the demand for information, the responsiveness of information search to the event occurs remarkably quickly: search behavior usually spikes on the same day as the event, and falls back to baseline within three or four days. This timing pattern suggests that Sims' (2003) mechanism for modelling rational attention more closely matches the data than the one proposed by Reis (2006).

We also present somewhat weaker evidence regarding to what extent acquiring information about taxes leads individuals to behave differently. We show that individuals sought information about capital losses following the 2008 stock market crash, especially leading up to the notch at the end of 2008, when harvesting losses provided an opportunity to reduce their tax liability. This is consistent with many investors not knowing the rules governing capital losses well enough to confidently apply them to make their behavior tax efficient in the wake of the crash. Information search is a necessary component of the response to tax incentives, especially where more obscure details of the tax code are concerned.

Future work should focus on how information-seeking affects behavior such as tax compliance, responses to tax enforcement, real responses such as labor supply, avoidance responses such as income shifting, as well as political discourse. This work will likely take advantage of an ever-expanding set of micro-data available about individual taxpayers' information search, both through official channels (official IRS resources), other commercial information searches (Lexis-Nexis, H&R Block), and non-profit providers of tax-related information (such as the Legal Information Institute at Cornell Law School). It may also take advantage of other methods of acquiring information generally (Twitter, Facebook, etc.). Future research should also take advantage of administrative and publicly available data to understand how individuals learn about other governmental policies (i.e., Social Security, unemployment insurance, healthcare, etc.).

Appendix

Obtaining and Analyzing Google Trends Data

When querying Google Trends, the user provides 1) search terms, 2) a geographical window, and 3) a time range. As we are studying a tax issue within the United States, for all data used in this paper, the geographical range is specified to be the United States.

When the number of overall searches for a given term is too low, Google Trends will report an SVI of zero, or will report daily at a weekly or monthly level (as opposed to daily SVI data). We encounter this issue if we query searches for "capital gains tax" alone. As such, we use a set of search terms to maximize our sample period for which we are able to obtain daily data. The search terms we include are, according to Google Trends itself, highly correlated with searches for "capital gains tax." The set of search terms is the following:

- Capital gains tax
- Capital gains tax rate
- Capital gains taxes
- Capital gains tax rate
- Capital gains calculator
- Capital gains
- Capital gains rate

We have verified that 1) the daily time series of SVI for simply "capital gains tax" is virtually identical to the one from the broader set of search terms, but with fewer missing data, and 2) we obtain nearly identical results for event studies and market movement effects if we use simply searches for "capital gains tax" as our left-hand-side variable instead of the broader set of search terms, but with slightly larger standard errors (reflecting the decreased number of observations).

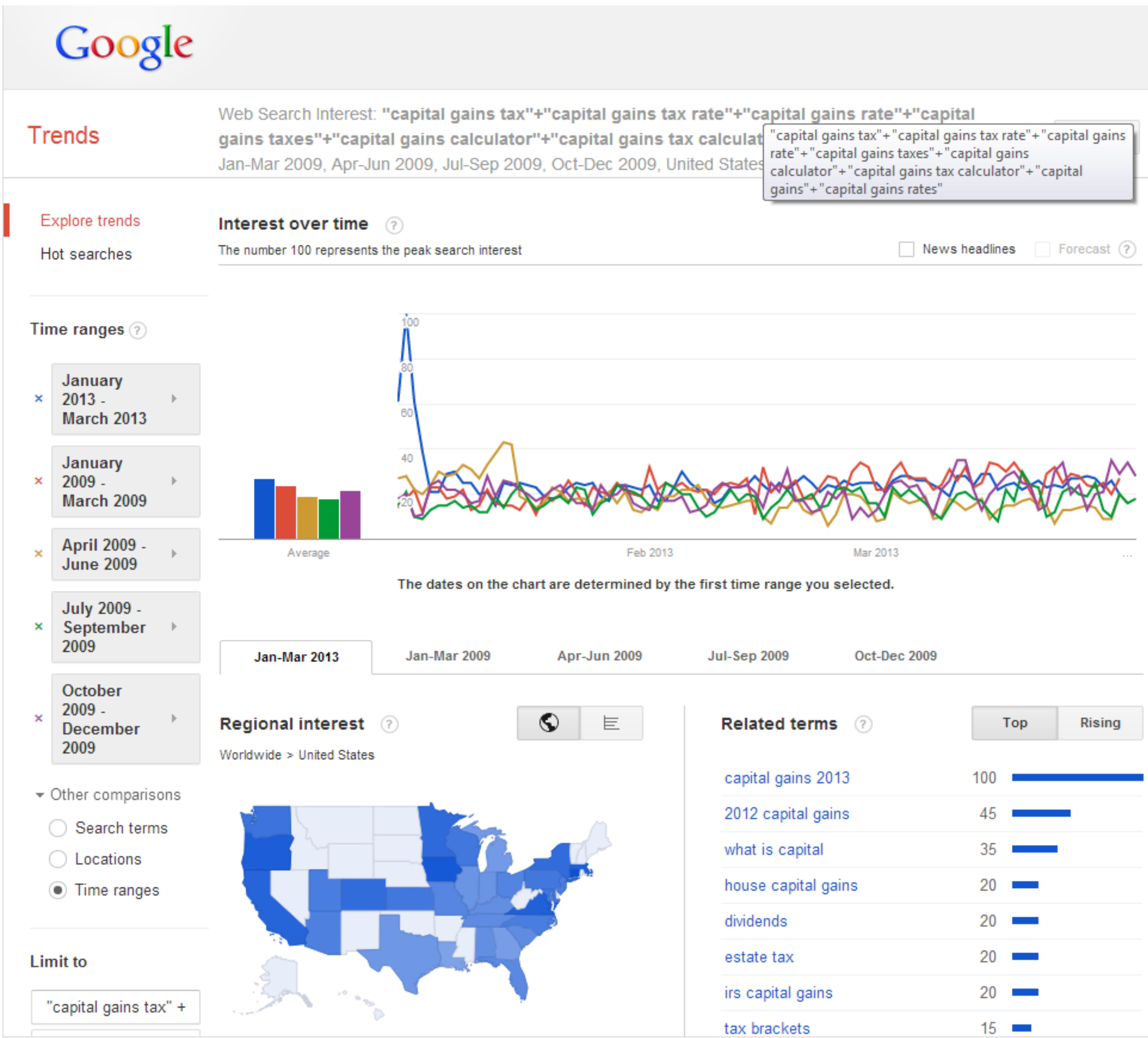
For the same reasons as above, we use multiple search terms related to stock advice. These are:

- Stock advice
- Stock market advice
- Stock tips
- Stocks to buy
- Stock to buy
- Stocks to sell
- Should buy stock
- Should sell stock
- Investment advice
- Investment tips

By default Google Trends provides weekly data when a user downloads a time series longer than three months. In order to access daily search volume data, one must query Google Trends in three month intervals. Fortunately, one can query several (up to five) three-month periods at once.

In order to obtain daily data while maintaining proper scaling of the variable across the entire time series, however, one must query Google Trends very carefully. The data are scaled so that SVI takes the value of 100 on the day with the highest propensity to search out of any date range and search terms. For example, if January 2, 2013 were the day with the highest propensity to search for capital gains tax terms (as it is), SVI would equal 100 on that day if January 2, 2013 were in the period provided by the user. In order to obtain daily data that is properly scaled for the full sample period, one must first find the single day with the highest search volume, and then include a time period containing that day along with sets of other three-month periods.

To get a properly scaled daily time series, we therefore include the time period January 1, 2013-March 31, 2013 in every single one of our queries for capital gains tax SVI, along with other three-month periods, until we obtain data for our entire sample period. The figure below shows an example of what such a query would look like to pull daily data for the year 2009. One can then download the data directly from this web page by clicking on the cog icon.



References

Askitas, N. and K. Zimmerman. (2009). "Google Econometrics and Unemployment Forecasting." *Applied Economics Quarterly* 55(2): 107-120.

Alstadsæter, A., W. Kopczuk, and K. Telle. (2010). Social Networks and Dynamics of Tax Avoidance: Evidence from a Well-defined Norwegian Tax Shelter. Working paper, Columbia University.

Auerbach, A.J. (1988). **"**Capital Gains Taxation in the United States: Realizations, Revenue, and Rhetoric." *Brookings Papers on Economic Activity* 2: 595-637.

Auerbach, A.J. (2002). "Taxation and Corporate Financial Policy." In *Handbook of Public Economics*, vol. 3, edited by Alan Auerbach and Martin Feldstein, pp. 1251-1292. North-Holland: Elsevier.

Blendon, R.J., S.R. Pelletier, M.D. Rosenbaum, and M. Brodie. (2003). "Tax Uncertainty." *Brookings Review* (Summer): 28-31.

Blumenthal, M. and J. Slemrod. (1992). "The Compliance Cost of the U.S. Individual Income Tax System: A Second Look after Tax Reform." *National Tax Journal* 45(2): 185-202.

Bronchetti, E.T., T.S. Dee, D.B. Huffman, and E. Magenheim. (2011). When a Nudge Isn't Enough: Defaults and Saving among Low-Income Tax Filers. National Bureau of Economic Research Working Paper No. w16887, Cambridge, MA.

Brown C.V. (1968). "Misconceptions About Income Tax and Incentives." *Scottish Journal of Political Economy* 16(2): 1-12.

Brownstein, J.S., C.C. Freifeld, and L.C. Madoff. (2009). "Digital Disease Detection—Harnessing the Web for Public Health Surveillance." *New England Journal of Medicine* 360(21): 2153-2157.

Burman, L. and W. Randolph. (1994). "Measuring Permanent Responses to Capital-Gains Tax Changes in Panel Data." *American Economic Review* 84(4): 794-809.

Chang, E. C., and J.M. Pinegar. (1986). Return Seasonality and Tax-loss Selling in the Market for Long-term Government and Corporate Bonds. *Journal of Financial Economics* 17(2): 391-415.

Chetty, R., A. Looney, and K. Kroft. (2007). "Salience and Taxation: Theory and Evidence." National Bureau of Economic Research Working Paper No. w13330, Cambridge, MA.

Chetty, R., A. Looney, and K. Kroft. (2009). "Salience and Taxation: Theory and Evidence." *American Economic Review* 99(4): 1145-1177.

Chetty, R. and E. Saez. (2009). Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients. National Bureau of Economic Research Working Paper No. w14836, Cambridge, MA.

Choi, H. and H. Varian. (2009). "Predicting Initial Claims for Unemployment." Working paper, Google Inc.

Choi, H. and H. Varian. (2012). "Predicting the Present with Google Trends." *Economic Record* 88(s1): 2-9.

Cloyd, C.B. and B.C. Spilker. (1999). "The Influence of Client Preferences on Tax Professionals' Search for Judicial Precedents, Subsequent Judgments and Recommendations." *The Accounting Review* 74(3): 299-322.

Connor, C.E., H.E. Egeth, and S. Yantis. (2004). "Visual Attention: Bottom-Up Versus Top-Down." *Current Biology* 14: R850-R852.

D'Auria, J.P. (2012). "Googling for Health Information." *Journal of Pediatric Health Care* 26(4):

e21-e23.
Da, Z., J. Engelberg, and P. Gao. (2011). “In Search of Attention.” *Journal of Finance* 66(5): 1461-1499.
Dai, Z., E. Maydew, D.A. Shackelford, and H.H. Zhang. (2008). Capital Gains Taxes and Asset Prices: Capitalization or Lock-in? *Journal of Finance* 63(2), 709-742.
de Bartoleme, C.A.M. (1995). “Which Tax Rate Do People Use: Average or Marginal?” *Journal of Public Economics* 56(1): 79-96.
Delli Carpini, M.X. and Keeter, S. (1996). *What Americans Know about Politics and Why It Matters*. New Haven, CT: Yale University Press.
Drake, M.S., D.T. Roulstone, and J.R. Thornock. (2012). "Investor Information Demand: Evidence from Google Searches around Earnings Announcements." *Journal of Accounting Research* 50(4): 1001-1040.
Eysenbach, G. (2006). “Infodemiology: Tracking Flu-related Searches on the Web for Syndromic Surveillance.” *AMIA Annual Symposium Proceedings Archive* 2006: 244-248.
Feldstein, M. and C. Clotfelter. (1976). "Tax Incentives and Charitable Contributions in the United States: A Microeconometric Analysis." *Journal of Public Economics* 5(1–2): 1-26.
Ginsberg, J., M.H. Mohebbi, R.S. Patel, et al. (2009). “Detecting Influenza Epidemics Using Search Engine Query Data.” *Nature* 457: 1012-1014.
Goel, S., J.M. Hofman, S. Lahaie, D.M. Pennock, and D.J. Watts. (2010). “Predicting Consumer Behavior with Web Search.” *Proceedings of the National Academy of Sciences of the United States of America* 107(41): 17486-17490.
Goolsbee, A. (2004). “The TurboTax Revolution: Can Technology Solve Tax Complexity?” In *The Crisis in Tax Administration*, edited by Henry J. Aaron and Joel Slemrod, pp. 124-38. Washington, DC: The Brookings Institution Press.
Gravelle, J.G. (1991). "Do Individual Retirement Accounts Increase Savings?" *Journal of Economic Perspectives*: 133-148.
Grossman, S. (1976). "On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information." *Journal of Finance* 31(2): 573-585.
Grossman, S. J. and J.E. Stiglitz. (1980). “On the Impossibility of Informationally Efficient Markets.” *American Economic Review* 70(3), 393-408.
Hanlon, D. and S. Pinder. (2012). Capital Gains Tax, Supply-driven Trading and Ownership Structure: Direct Evidence of the Lock-in Effect. *Accounting and Finance* 53(2):419-439.
Hultha, A., G. Rydevik, and A. Linde. (2009). “Web Queries as a Source for Syndromic Surveillance.” *PLOS ONE* 4(2): e4378.
Ivković, Z., J. Poterba, and S. Weisbenner (2005). “Tax-Motivated Trading by Individual Investors.” *American Economic Review* 95(5): 1605-1630.
Johnson, H.A., M.M. Wagner, W.R. Hogan, W. Chapman, R.T. Olszewski et al. (2004). “Analysis of Web Access Logs for Surveillance of Influenza.” *Studies in Health Technology and Informatics* 107(2): 1202-1206.
Kharkar, G. (2007). "Googling for Help: Lawyer Referral Services and the Internet." *Georgetown Journal of Legal Ethics* 20(3): 769-778.
Kopczuk, W. and J. Slemrod. (2003). “Dying to Save Taxes: Evidence from Estate-Tax Returns on the Death Elasticity.” *Review of Economics and Statistics* 85(2), 256-265.
Laurent, M.R. and T.J. Vickers. (2009). "Seeking Health Information Online: Does Wikipedia Matter?" *Journal of the American Medical Informatics Association* 16(4): 471-479.

Liebman, J.B. and E.F.P. Luttmer. (2011). Would People Behave Differently If They Better Understood Social Security? Evidence from a Field Experiment. National Bureau of Economic Research Working Paper No. w17287, Cambridge, MA.

Maćkowiak, B. and M. Wiederholt. (2009). "Optimal Sticky Prices under Rational Inattention." *American Economic Review* 99(3) 769-803.

Mestyán, M., T. Yasseri, and J. Kertész. (2012). Early Prediction of Movie Box Office Success based on Wikipedia Activity Big Data. arXiv:1211.0970v2 [physics.soc-ph].

Meyer, B.D. and D.T. Rosenbaum (2000). "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." *Quarterly Journal of Economics* 116(3): 1063-1114.

Mohebbi, M., D. Vanderkam, J. Kodysh, R. Shonberger, H. Choi, et al. (2011). "Google Correlate Whitepaper." Google Inc.

Muth, J. F. (1961). "Rational Expectations and the Theory of Price Movements." *Econometrica* 29(3): 315-335.

Odean, T. (1998). "Are Investors Reluctant to Realize their Losses?" *Journal of Finance 53*(5): 1775-1798.

Pelat, C., C. Turbelin, A. Bar-Hen, A. Flahault, and A-J. Valleron. (2009). "More Diseases Tracked by Using Google Trends." *Emerging Infectious Diseases* 15(8):1327-1328.

Polgreen P.M., Y. Chen, D.M. Pennock, and F.D. Nelson (2008). "Using Internet Searches for Influenza Surveillance." *Clinical Infectious Diseases* 47(11): 1443-1448.

Poterba, J.M. (1987). "How Burdensome Are Capital Gains Taxes? Evidence from the United States." *Journal of Public Economics* 33(2): 157-172.

Poterba, J.M. and S.J. Weisbenner. (2001). "Capital Gains Tax Rules, Tax Loss Trading, and Turn-of-the-Year Returns," *Journal of Finance* 56(1): 353-368.

Ivković, Z., Poterba, J., and S.J. Weisbenner. (2005). Tax-Motivated Trading by Individual Investors. *American Economic Review 95*(5), 1605-1630.

Poterba, J.M. (2001)."Estate and Gift Taxes and Incentives for Inter Vivos Giving in the United States." *Journal of Public Economics* 79(1): 237-264.

Prendiville, T.W., J. Saunders, and J. Fitzsimons. (2009). "The Information-Seeking Behaviour of Paediatricians Accessing Web-Based Resources." *Archives of Disease in Childhood* 94(8): 633-635.

Recupero, P.R., S.E. Harms, and J.M. Noble. (2008). "Googling Suicide: Surfing for Suicide Information on the Internet." *Journal of Clinical Psychiatry* 69(6):878-888.

Sahm, C.R., M.D. Shapiro, and J. Slemrod. (2012). "Check in the Mail or More in the Paycheck: Does the Effectiveness of Fiscal Stimulus Depend on How It Is Delivered?" *American Economic Journal: Economic Policy* 4(3): 216–250.

Schmidt, T. and S. Vosen. (2011). "Forecasting Private Consumption: Survey-based Indicators vs. Google Trends." *Journal of Forecasting* 30(6): 565-578.

Sheffrin, S.M. (1994). "Perceptions of Fairness in the Crucible of Tax Policy." In *Tax Progressivity and Income Inequality*, edited by Joel Slemrod, pp. 309-334. Cambridge: Cambridge University Press.

Sims, C. A. (2003). "Implications of Rational Inattention." *Journal of Monetary Economics* 50(3): 665-690.

Sims, C. A. (2006). "Rational Attention: Beyond the Linear Quadratic Case." *The American Economic Review* 96(2): 158-163.

Slemrod, J. (1992). "Did the Tax Reform Act of 1986 Simplify Tax Matters?" *Journal of Economic Perspectives* (Winter): 45-57.

Slemrod, J. (1995). “The Compliance Cost of Income Taxation, and How to Reduce It,” National Commission on Economic Growth and Tax Reform (Kemp Commission), Oral testimony given July 12, 1995.

Slemrod, J. (2004). “New Estimates of the Compliance Cost of Income Taxation,” Committee on Ways and Means, Subcommittee on Oversight, Hearing on Tax Simplification, invited written testimony given June 15, 2004.

Slemrod, J. and N. Sorum. (1985). “The Compliance Cost of the U.S. Individual Income Tax System.” *National Tax Journal* 37(4): 461-474.

Speckman, P. (1986). “Kernel Smoothing in Partial Linear Models.” *Journal of the Royal Statistical Society* 50(3): 413-436.

Starks, L. T., L. Yong, and L. Zheng. (2006). “Tax-Loss Selling and the January Effect: Evidence from Municipal Bond Closed-End Funds.” *Journal of Finance* 61(6): 3049-3067.

Stiglitz, J.E. (1983). "Some Aspects of the Taxation of Capital Gains." *Journal of Public Economics* 21(2): 257-294.

Tang, H. and J.H.K. Ng. (2006). "Googling for a Diagnosis—Use of Google as a Diagnostic Aid: Internet Based Study." *BMJ* 333(7579): 1143-1145.

Theeuwes, J. (1991). “Exogenous and Endogenous Control of Attention: The Effect of Visual Onsets and Offsets.” *Perception & Psychophysics* 49(1): 83-90.

Wilson, K., and J.S. Brownstein. (2009). “Early Detection of Disease Outbreaks Using the Internet.” *Canadian Medical Association Journal* 180(8): 829-831.

Wheeler, R.E. (2011). "Does Westlawnext Really Change Everything? The Implications of Westlawnext on Legal Research." *Law Library Journal* 103(3): 359-377.

Xiang, Z. and U. Gretzel. (2010). "Role of Social Media in Online Travel Information Search." *Tourism Management* 31(2): 179-188.

Yantis, S. (2000). “Goal-directed and Stimulus-driven Determinants of Attentional Control.” *Attention and Performance* 18: 73-103.

Table I: Testable Implications of Theories of Information and Attention

Behavioral assumption	Should information search respond to… time notches?	macroeconomic change?	enactment of policy changes?	implementation of policy changes?	filing deadlines?	news events?
Full information	No	No	No	No	No	No
Rational expectations	No	No	Yes, instantly	No	No	No
Rational attention, Sims (2003)	Yes	Yes, instantly	Yes, instantly	Yes, instantly	Yes	No*
Rational attention, Reis (2006)	Yes	Yes, with delay**	Yes, with delay**	Yes, with delay**	Yes	No*
Exogenous Attention (Salience)	No	No	Yes, while in the news	Yes, while in the news	Yes	Yes
Summary of Our Findings	Yes: Dec 31 time notch, esp. capital losses (due to loss harvesting)	Yes, instantly: day-of stock market trading volume, response to stock market crash	Yes, instantly/while in the news: ATRA, presidential debates and elections	Yes: May 2008 stimulus	Yes: seasonality due to tax season	Yes, instantly: release of Mitt Romney's tax returns

Notes: *News events may inspire attention “rationally" due to a preference for being informed in public discussion of current events, but it will not inspire rational attention for the purpose of economic behavior, either capital gains realizations or tax compliance. ** By "with delay," we mean that the surge in information search as a result of the event should be spread out over a longer time period in the Reis (2006) model compared to the Sims (2003) model, as different taxpayers reach the date at which they (previously) decided to update at different times.

Table II: Events affecting Taxpayer Information Search

Date	Description
Oct 15, 2008*	Presidential debate between Barack Obama and John McCain. Candidates discuss their capital gains tax proposals.
November 4, 2008	Barack Obama elected President of the United States
January 24, 2012	Mitt Romney releases his 2010 tax return
September 21, 2012	Mitt Romney releases his 2011 tax return
Oct 16, 2012	Presidential debate between Barack Obama and Mitt Romney. Candidates discuss Romney's capital gains tax proposals.
Nov 6, 2012	Barack Obama re-Elected
Jan 2, 2013	American Taxpayer Relief Act signed into law, includes an increase in top capital gains tax rate

Notes: ***This was also the date of a large movement in the stock market. We discuss the difficulty distinguishing the stock market crash from political events in October of 2008 in Section 6. See also Figure 5.

Table III: Estimating the Impact of Events on Information Search for Capital Gains Taxes

	Google Searches	Wikipedia Page Views	Calls to IRS
Obama/McCain Debate	12.08	5.469	-0.222
	93.893	2662.510	-16.171
	(0.000)	(0.015)	(1.000)
Obama Elected	16.04	5.873	-0.132
	124.658	2914.756	-12.105
	(0.000)	(0.008)	(0.999)
Mitt Romney's 2010 tax return released	20.820	34.959	0.379
	161.000	17351.698	-559.084
	(0.000)	(0.000)	(0.000)
Mitt Romney's 2011 tax return released	20.720	12.647	-0.341
	55.689	6277.817	-32.471
	(0.004)	(0.000)	(0.999)
Obama/Romney Debate	9.39	9.880	0.251
	72.989	4904.106	-17.693
	(0.000)	(0.000)	(0.999)
Obama Re-Elected	9.14	14.563	-0.31
	71.023	7228.503	-14.267
	(0.000)	(0.000)	(1.000)
American Taxpayer Relief Act signed	15.310	9.623	-6.170
	118.978	4776.242	-188.522
	(0.000)	(0.000)	(0.000)
Sample Period	Jan 2, 2004— March 21, 2013	Dec 9, 2007— March 31, 2013	Jan 2, 2002— March 31, 2013
Number of Days	3364	1938	4100

Notes: For each event, the top number reports the cumulative information search attributed to this event, in daily standard deviation units, added over the five-day event window. The second number reports the same estimate in the original units of the search volume measure, i.e. the Google Trends index or, Wikipedia page views, or number of calls to the IRS. The bottom number, in parenthesis, reports the p-value from an F-test in which the null hypothesis is that the event had no impact on information search, i.e. that the variation in information searches over the event window is purely random. The estimation controls for variation due to yearly and weekly seasonality.

Table IV. Information Search and Stock Market Activity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample Period	Jan 2, 2004— Dec 29, 2012	Dec 9, 2007— Dec 29, 2012	Jan 7, 2002— Dec 29, 2012	Sep 1, 2008— Feb 28, 2009	Sep 1, 2008— Feb 28, 2009	Jan 2, 2004— Dec 29, 2012, except Sep 1, 2008— Feb 28, 2009	Dec 9, 2007— Dec 29, 2012, except Sep 1, 2008— Feb 28, 2009
Dependent Variable:	Google Searches (Standardized)	Wikipedia Page Views (Standardized)	Calls to IRS (Standardized)	Google Searches (Standardized)	Wikipedia Page Views (Standardized)	Google Searches (Standardized)	Wikipedia Page Views (Standardized)
Log Trading Volume	0.286	0.306	0.043	1.071	0.926	0.215	0.386
	(0.156)*	(0.159)*	(0.075)	(0.780)	(0.470)*	(0.159)	(0.171)**
Log Trading Volume Lag 1	-0.051	-0.072	0.099	-0.141	0.263	-0.11	-0.093
	(0.146)	(0.151)	(0.070)	(0.708)	(0.415)	(0.149)	(0.162)
Log Trading Volume Lead 1	0.062	-0.023	-0.037	-0.125	-0.123	0.11	-0.05
	(0.177)	(0.181)	(0.085)	(0.887)	(0.553)	(0.179)	(0.193)
Number of Days	2252	1266	2753	124	124	2128	1142

Notes: The dependent variable and log trading volume are de-trended prior to estimation, and the dependent variable is standardized by dividing by the standard deviation of the detrended data. We control for weekly and yearly seasonality and the events in Table 2 (except those events that occur outside the sample period, since market data are not available past December 29, 2012). The regression included five market-dated leads and lags of each measure. We only report one lead and one lag for brevity and clarity. The coefficients are similar if we use 14 leads and lags instead of 5, and none of the results change substantially with the inclusion of lags of the dependent variable. Standard errors are provided in parentheses below point estimates. * indicates p<0.10, ** indicates p<0.05, and *** indicates p<0.01. The standard deviation of log trading volume is 0.223.

Table V: Searches for Stock Advice and Searches for Capital Gains Tax

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample Period	Jan 2, 2004— Dec 29, 2012	Jan 2, 2004— Dec 29, 2012	Jan 7, 2004— Dec 29, 2012	Jan 7, 2004— Mar 21, 2013	Dec 9, 2007— Dec 29, 2012	Dec 9, 2007— Dec 29, 2012	Dec 9, 2007— Dec 29, 2012	Dec 9, 2007— Mar 31, 2013
Dependent Variable:	Google Searches (Standardized)				Wikipedia Page Views (Standardized)			
Log Trading Volume	0.286	0.174	0.130		0.306	0.162	0.178	
	(0.156)*	(0.155)	(0.155)		(0.159)*	(0.157)	(0.157)	
Log Trading Volume Lag 1	-0.051	-0.100	-0.06		-0.072	-0.141	-0.137	
	(0.146)	(0.145)	(0.146)		(0.151)	(0.147)	(0.148)	
Log Trading Volume Lead 1	0.062	0.023	0.002		-0.023	-0.056	-0.098	
	(0.177)	(0.175)	(0.175)		(0.181)	(0.177)	(0.178)	
Stock Advice Search		0.031	0.008	0.009		0.030	0.016	0.017
		(0.004)***	(0.005)	(0.005)*		(0.004)***	(0.006)***	(0.005)***
Stock Advice Search Lag 1			0.017	0.017			0.017	0.015
			(0.006)***	(0.006)***			(0.008)**	(0.008)*
Stock Advice Search Lead			0.009	0.008			0.005	0.003
			(0.005)**	(0.005)*			(0.006)	(0.006)
Number of Days	2252	2239	2200	2211	1266	1266	1266	1272

Notes: The dependent variable, log trading volume, and stock advice search volume are de-trended prior to estimation, and the dependent variable is standardized by dividing by the standard deviation of the detrended data. We control for weekly and yearly seasonality and the events in Table 1 (except those events that occur outside the sample period, since market data are not available past December 29, 2012). The regression included five market-dated leads and lags of trading volume and stock advice search volume. We only report one lead and one lag for brevity and clarity. Standard errors are provided in parentheses below point estimates. * indicates p<0.10, ** indicates p<0.05, and *** indicates p<0.01.

The standard deviation of log trading volume is 0.223. The standard deviation of Stock advice search volume is 4.458.

Table VI: Delayed Effects of Search on Market Outcomes

	Google Searches (Standardized)	Wikipedia Page Views (Standardized)	Calls to IRS (Standardized)
Log Trading Volume, This Year	0.083	-0.169	-0.005
	(0.136)	(0.150)	(0.050)
Log Trading Volume, Previous Year	-0.039	0.263	-0.011
	(0.107)	(0.268)	(0.046)
End of Year *x* Log Volume This Year	0.029	0.036	-0.004
	(0.006)***	(0.006)***	(0.003)
Tax Season *x* Log Volume Previous Year	0.028	0.001	0.067
	(0.007)***	(0.007)	(0.003)***
Market Return, This Year	-0.161	-0.227	-0.015
	(0.111)	(0.206)	(0.037)
Market Return, Previous Year	0.094	0.01	0.024
	(0.100)	(0.121)	(0.248)
End of Year *x* Market Return This Year	1.502	1.975	0.238
	(0.606)**	(0.514)***	(0.248)
Tax Season *x* Market Return Previous Year	0.149	0.467	-0.142
	(0.165)	(0.156)***	(0.072)
Sample Period	Jan 2, 2004— Dec 31, 2012	Dec 9, 2007— Dec 31, 2012	Jan 7, 2002— Dec 31, 2012
Number of Days	3025	1695	3684

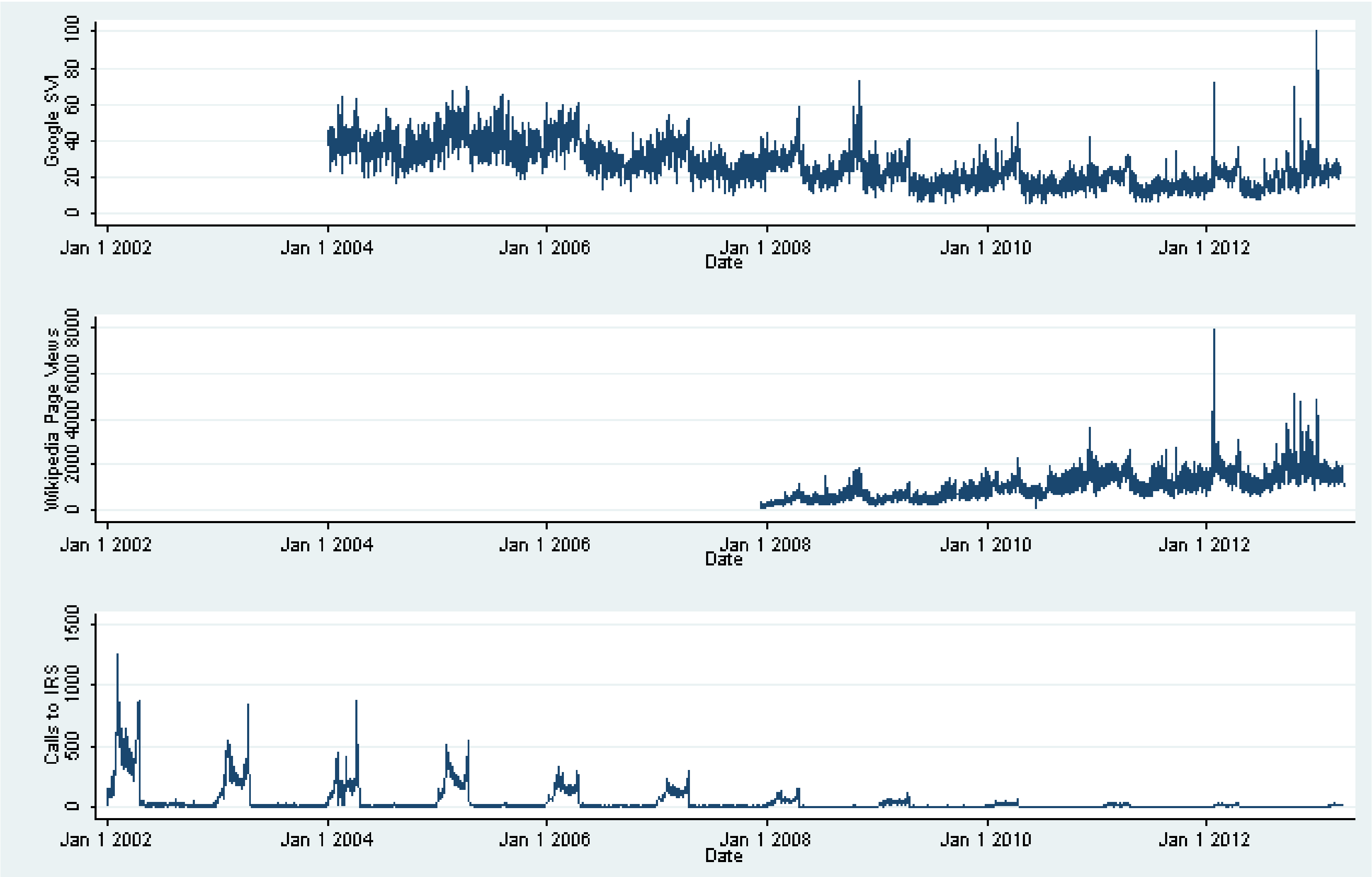
Notes: Data are de-trended prior to estimation, and the dependent variable is standardized by dividing by the standard deviation. We control for weekly and yearly seasonality and the events in Table 2 (except those events that occur outside the sample period, since market data are not available past Apr 8, 2012).
Standard errors are provided in parentheses below point estimates. * indicates p<0.10, ** indicates p<0.05, and *** indicates p<0.01.

Table VII: Most Popular External Referring Sites

Rank	External Referring Site	Visits
1	http://www.google.com/	77,806,234
2	Direct Traffic	21,134,463
3	http://www.bing.com/	12,547,034
4	http://search.yahoo.com/	11,151,192
5	http://36ohk6dgmcd1n-c.c.yom.mail.yahoo.net/	2,871,475
6	http://www.irs.gov/	1,543,764
7	http://search.aol.com/	1,406,364
8	http://www.hrblock.com/	1,369,328
9	http://www.ask.com/	1,281,277
10	http://turbotax.intuit.com/	888,581
11	http://taxes.about.com/	726,794
12	http://search.comcast.net/	587,255
13	http://search.mywebsearch.com/	526,147
14	http://links.govdelivery.com/	489,629
15	https://www.google.com/	487,301
16	http://www.search-results.com/	434,511
17	http://isearch.avg.com/	373,820
18	http://mail.aol.com/	370,012
19	http://www.google.ca/	354,293
20	http://www.google.co.uk/	249,716
21	http://m.yahoo.com/	247,181
22	http://apps.irs.gov/	239,659
23	http://www.google.co.in/	238,667
24	http://search.irs.gov/	207,177
25	http://us.mg5.mail.yahoo.com/	200,605
26	http://answers.yahoo.com/	197,800
27	http://irweb.irs.gov/	188,232
28	http://finance.yahoo.com/	183,980
29	http://us.mg4.mail.yahoo.com/	180,243
30	http://www.ehow.com/	175,433
31	http://m.bing.com/	173,380
32	http://www.google.com.pr/	149,668
33	http://search.babylon.com/	148,311
34	http://file.gov.com/	134,832
35	http://en.wikipedia.org/	116,423
36	http://www.facebook.com/	112,541
37	http://www.sba.gov/	109,391
38	http://us.yhs4.search.yahoo.com/	109,172
39	http://www.webcrawler.com/	108,954
40	http://search.rr.com/	103,668

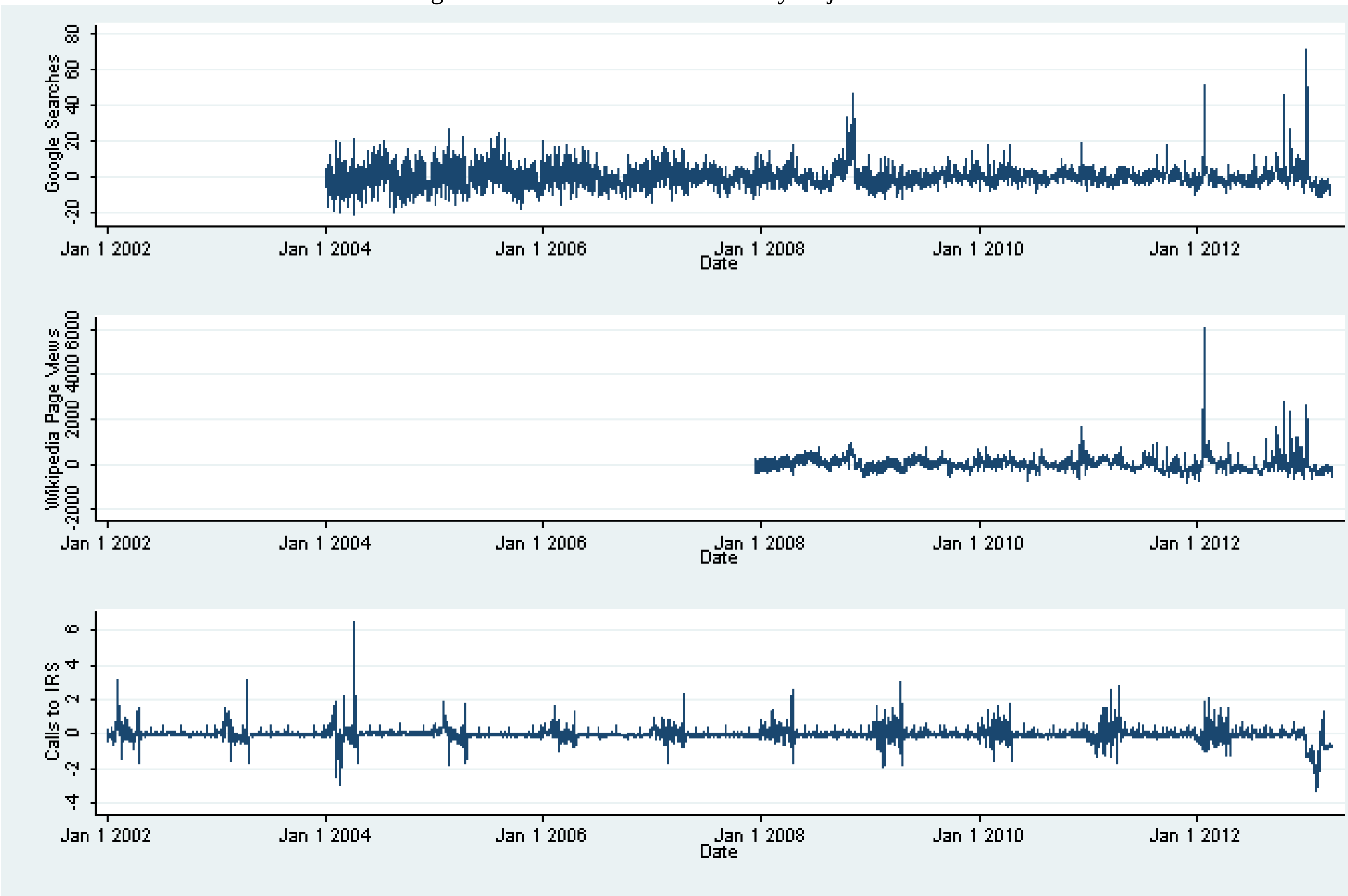
Notes: These are the number of visitors to the IRS.gov domain that are refereed from each of the above sites, from November 2011-April 2012.

Figure I: Plots of the Raw Data



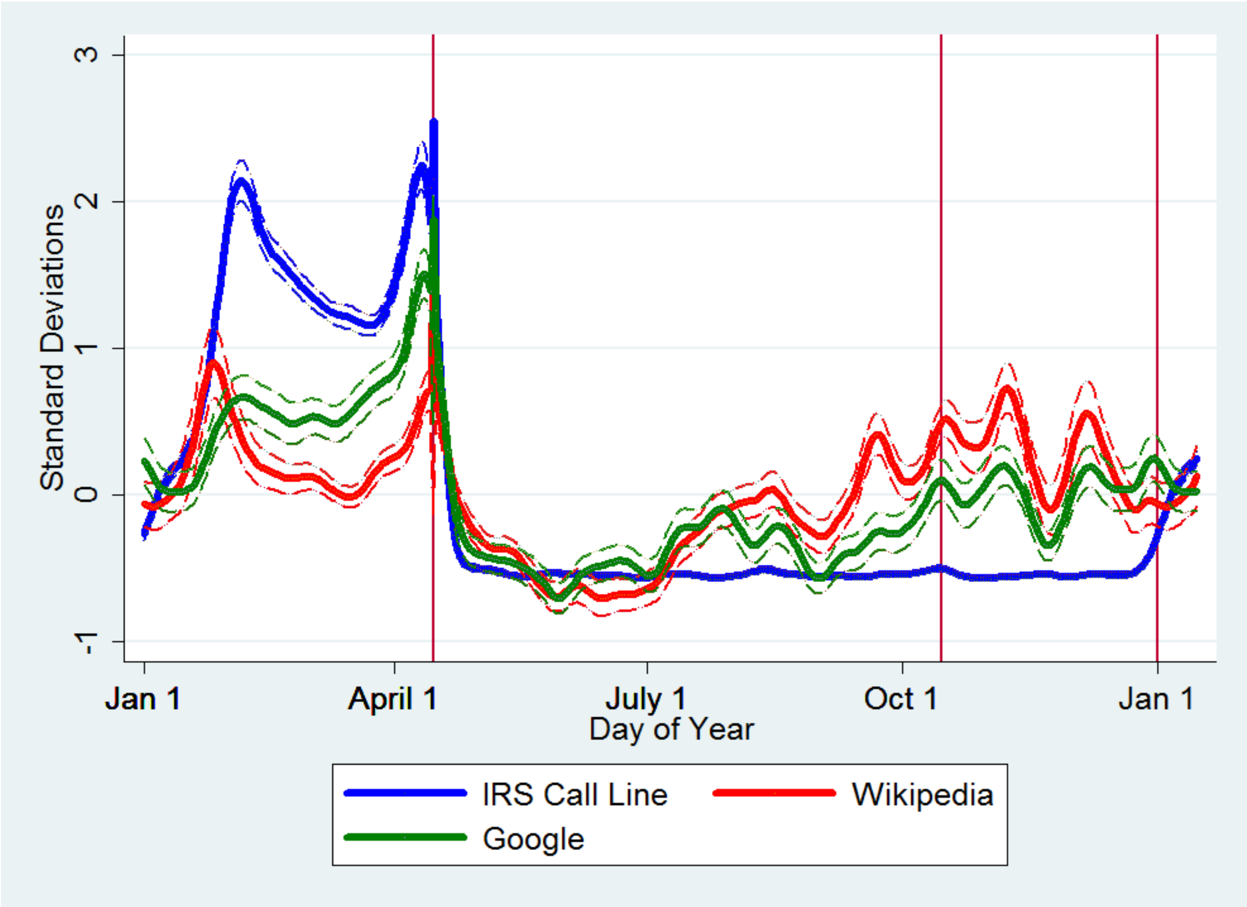
Notes: These graphs show the raw data for (in order) Google search volume, Wikipedia page views, and calls made to the IRS.

Figure II: Detrended and Seasonally Adjusted Data



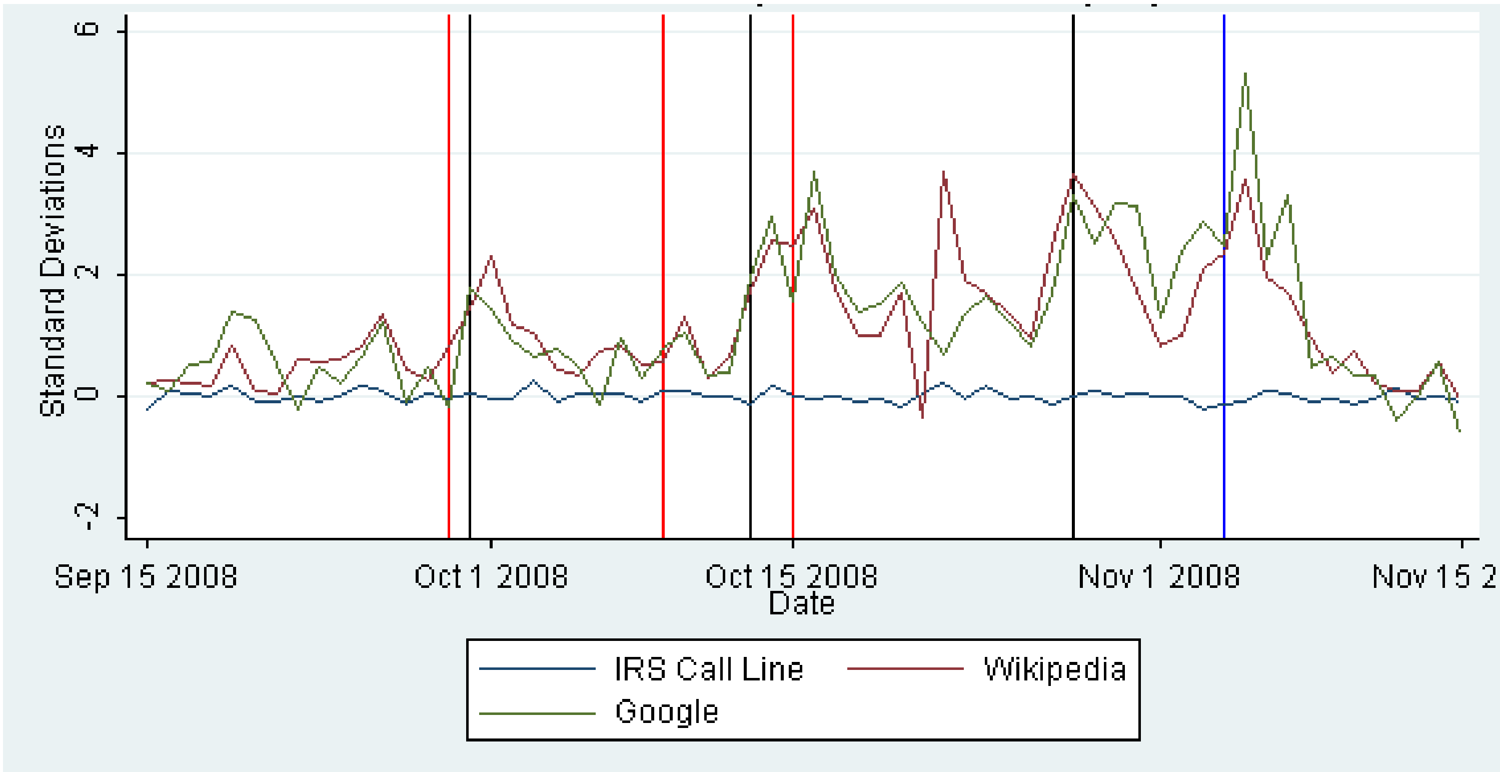
Notes: These graphs show the detrended data for Google search volume, Wikipedia page views, and calls made to the IRS.

Figure III: Yearly Seasonality



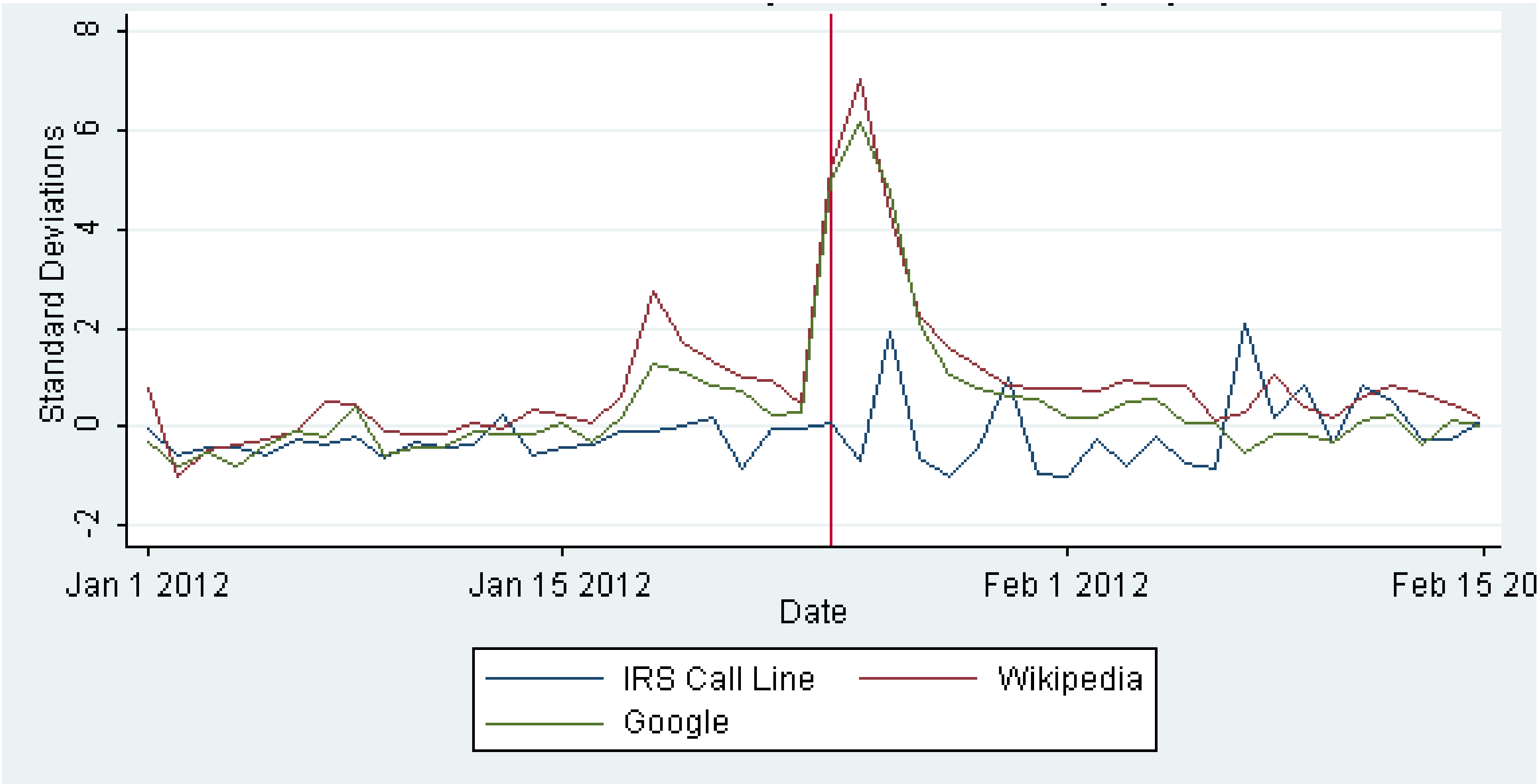
Notes: Yearly seasonality is estimated using the smooth-fixed-effects method described in Section 5, with controls for the events in Table 2 and the market movements in Table 4. The dashed lines represent the bounds of 95 percent confidence intervals. The three vertical lines correspond to April 15, October 15, and January 1, respectively.

Figure IV: Capital Gains Tax Information Search in October 2008



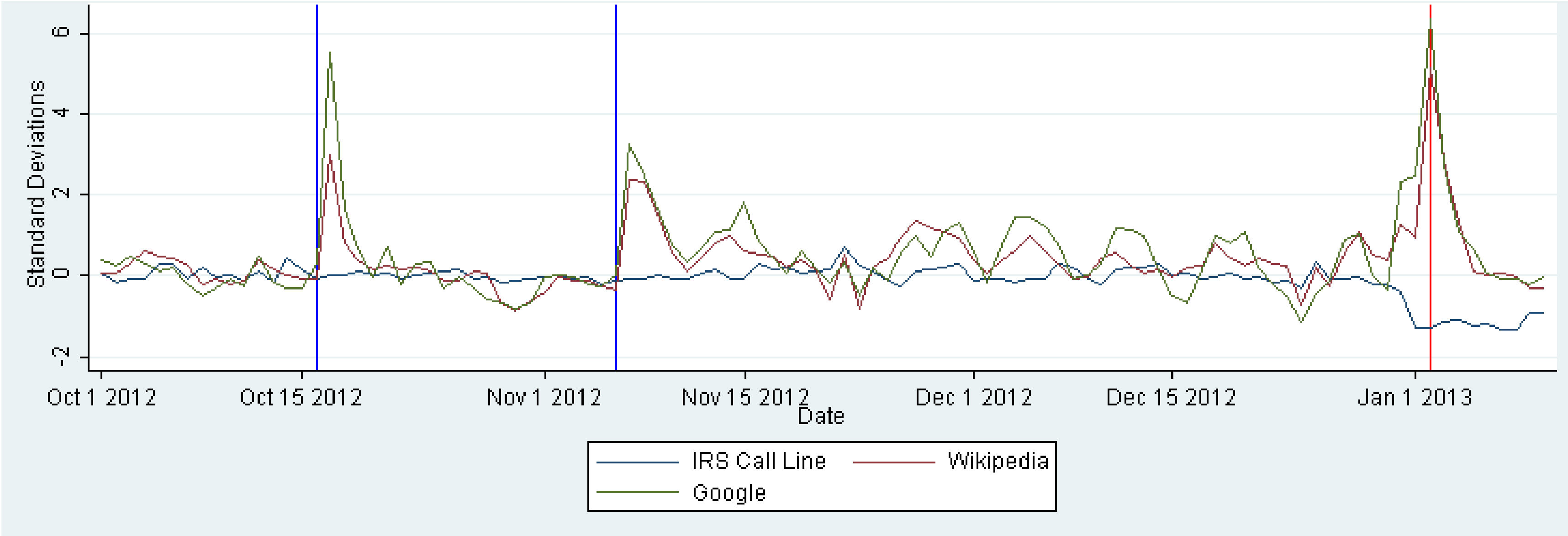
Notes: We plot the detrended, standardized, and seasonally adjusted data (from Figure 2) over a narrower time period to examine the spike in October 2008. The vertical red lines mark the worst three days of the stock market crash of 2008 according to the percent drop in the S&P 500 index: October 15 (-9.03%) , September 29 (-8.81%), and October 9 (-7.62%). The black lines mark the three largest gains in the S&P500 during this period of extreme volatility: October 13 (11.58%), October 28 (10.79%) and September 30 (5.42%). October 15, 2008 was also the date of a Presidential Debate between Barack Obama and John McCain, in the course of which both candidates made proposals for changing capital gains tax rates (The other debates occurred on September 26 and October 7, and did not discuss capital gains taxes). The vertical blue line marks the date of the 2008 Presidential election, November 4, 2008.

Figure V: Capital Gains Tax Information Search in January 2012



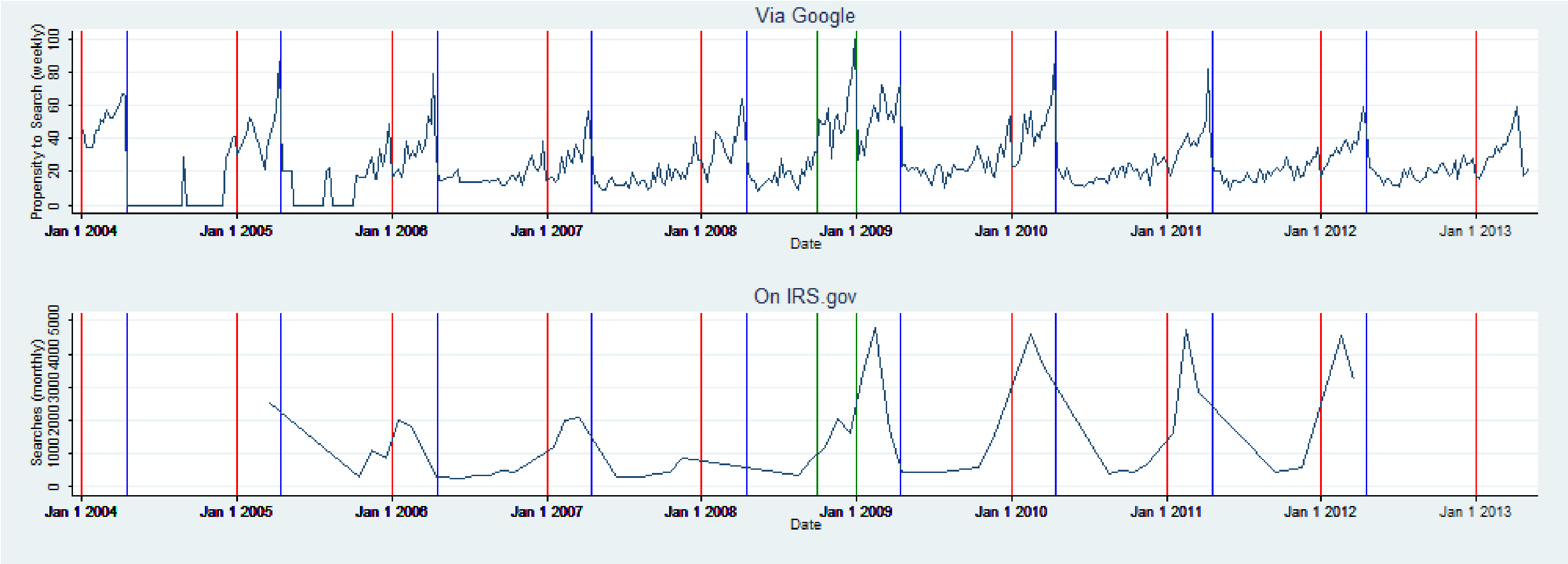
Notes: We plot the detrended, standardized, and seasonally adjusted data (from Figure 4) over a narrower time period to examine the spike in January 2012. The vertical red line marks January 24, 2012, the date that Mitt Romney released his 2010 tax return.

Figure VI: Capital Gains Tax Information Search in November 2012 – January 2013



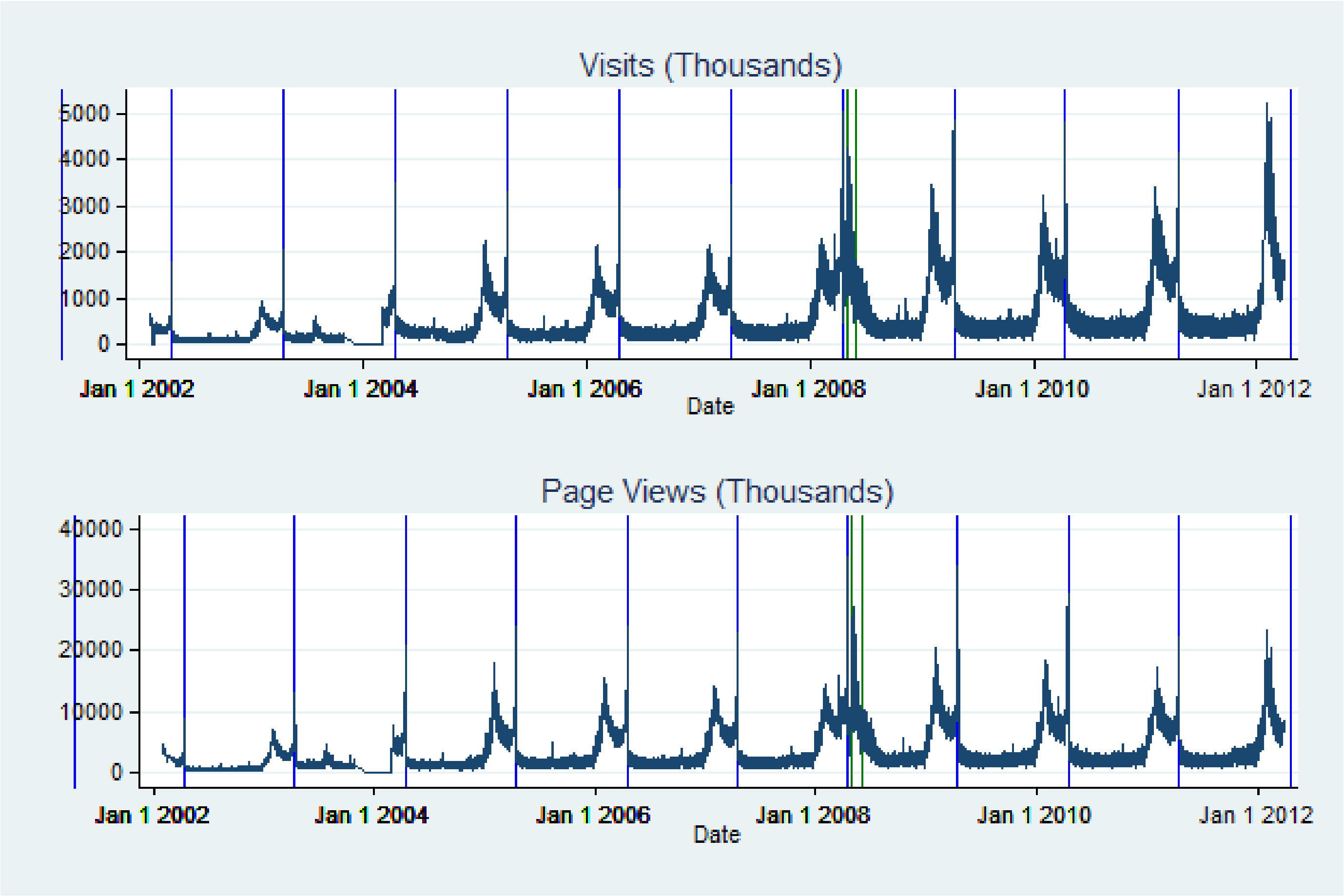
Notes: The first blue line marks the date of the Presidential debate on October 16, 2012, during which Barack Obama and Mitt Romney debated the merits of Romney's proposals for lowering capital gains tax rates. The second blue line marks the date of the Presidential election, November 7, 2013. The red vertical line marks the passage of the American Taxpayer Relief Act on January 2, 2013.

Figure VII: Information Search on Capital Losses



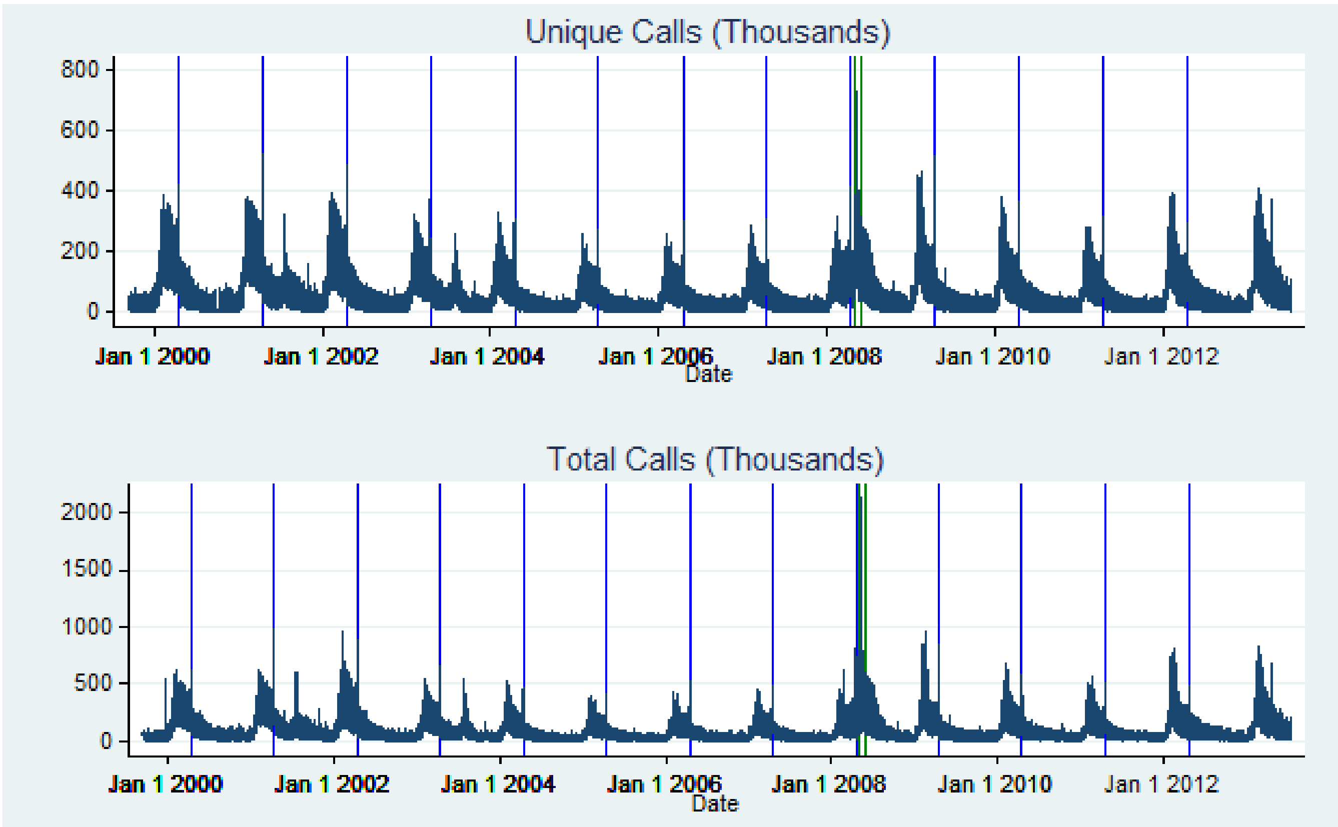
Notes: The first panel plots weekly Google Trends data on searches for just "capital loss." Google Trends data are missing for some dates in 2004 and 2005, when search volume was too low and Google does not provide that data. The second graph plots monthly searches in the search bar on the IRS home page for terms related to capital losses. The green lines correspond to October 1 and December 2008, to delineate the period in which there was increased information search on capital losses due to the 2008 stock market crash. Red lines denote January 1st of each year, to highlight that a spike in search volume typically occurs at the very end of each year. Blue lines denote mid-April filing deadlines, to highlight the spike occurring in the run-up to the filing deadline.

Figure VIII: Visits and Page Views of all URLs on the IRS.gov domain.



Notes: Data are unavailable from November 1, 2003 to February 29, 2004. The green vertical lines delineate the month of May 2008. The blue vertical lines denote the mid-April filing deadline for each year.

Figure IX: All Calls to IRS *Individual Assistance Call Line*



Notes: Unique callers are identified by the telephone number from which the call originates. The green vertical lines delineate May 2008. The blue vertical lines denote the mid-April filing deadline for each year.